An evolutionarily stable strategy to colonize spatially extended habitats

The ability of a species to colonize newly available habitats is crucial to its overall fitness. In general, motility and fast expansion are expected to be beneficial for colonization and hence for the fitness of an organism. Here we apply an evolution protocol to investigate phenotypical requirements for colonizing habitats of different sizes during range expansion by chemotaxing bacteria. Contrary to the intuitive expectation that faster is better, we show that there is an optimal expansion speed for a given habitat size. Our analysis showed that this effect arises from interactions among pioneering cells at the front of the expanding population, and revealed a simple, evolutionarily stable strategy for colonizing a habitat of a specific size: to expand at a speed given by the product of the growth rate and the habitat size. These results illustrate stability-to-invasion as a powerful principle for the selection of phenotypes in complex ecological processes.

When an organism encounters an unoccupied habitat, it colonizes the habitat through growth and expansion. A recent quantitative study of chemotaxis-mediated bacterial range expansion showed that the characteristics of the expanding population are dominated by a group of pioneering cells at the population front; these pioneers move outwards, replicate, and leave behind offspring (settlers) to grow and occupy the territories traversed. To understand the determinants of colonization behind the front, we modified common experimental evolution protocols to select for cells at various distances from the point of the initial invasion, well after the passage of the front. Motile Escherichia coli cells inoculated at the centre of a semi-solid tryptone broth (TB) agar plate were given time to expand outwards and colonize the entire plate. Twenty-four hours after inoculation, after the entire plate had been filled with bacteria, a small volume of agar (containing about 7.4 × 10^6 cells) was taken at a specific radius and transferred directly to the centre of a fresh plate. The process was repeated for 50 cycles (about 600 generations), with the transferred samples always taken at the same distance from the centre. Five such series were generated at five different distances from the origin (positions A–E in Extended Data Fig. 1c).

We first evaluated the growth and expansion characteristics of samples collected at various cycles and positions. When inoculated on fresh TB plates, the evolved populations still expanded steadily outwards (Extended Data Fig. 1d), but were characterized by expansion speeds with distinct dependencies on the selection distance (Fig. 1b). Populations collected from outer radii exhibited a steady increase in expansion speeds, whereas those collected from inner radii exhibited a steady decrease, leaving an intermediate selection distance (position C, with distance Xₐ = 15 mm) at which the expansion speed of the evolved populations fluctuated around that of the ancestor (about 6 mm h⁻¹) throughout the evolution process. These results were highly reproducible across replicates and in different growth media containing generic amino acid supplements (Extended Data Fig. 2a–c). The divergent pattern of the evolved expansion speeds obtained (Fig. 1b) is not the result of a simple trade-off with the rate of cell growth, as the batch culture growth rates changed very little for strains evolved in TB (Extended Data Fig. 1e). Strains evolved in casamino acids (CAA) generally showed growth rates higher than the ancestor (Extended Data Fig. 1f, g); however, changes in their expansion speeds still followed the divergent pattern according to their selection positions (Extended Data Fig. 2c). By contrast, expansion speeds increased regardless of selection position for strains evolved in medium with glycerol and no chemoattractants (Extended Data Fig. 2d), suggesting that chemotaxis is important for the divergent evolution phenomenon shown in Fig. 1.

An examination of 300 strains sampled from the 50th cycle of the five evolved series across three replicates showed that the distributions of expansion speeds of individual strains were well reflected by the previous measurements of samples containing mixed populations (Fig. 1c). Furthermore, changes in expansion speeds were consistent with changes in the motility characteristics of the evolved cells obtained from single-cell analysis (Extended Data Fig. 1h, i), and with mutations identified from genomic sequence analysis. Sequencing of population samples at the 50th cycle of each evolved series yielded a multitude of mutations (Supplementary Table 1). Several dominant mutations were introduced individually into the ancestral strain; these were found to change the expansion speeds of the ancestral strain towards those of the evolved strains from which the mutations were derived (Extended Data Fig. 1j).

Two-strain competition in space

To understand the underlying evolutionary process, we first compared the expansion dynamics of clonal populations that were grown...
individually. We chose several mutant strains isolated at the 50th cycle that exhibited a range of expansion speeds but had similar growth rates (A, B, B’, C, D, and E in Fig. 1c; see Supplementary Table 2). We transformed each strain with GFP and calibrated their fluorescence intensities by direct cell counting (Extended Data Fig. 3a–c). This allowed direct observations of the spatiotemporal dynamics of density profiles of each strain (Extended Data Fig. 4a–c for ancestor, mutant B, and mutant D). Clearly, faster strains showed higher abundances at all positions and all times.

Next, we competed each mutant strain against the ancestor strain. We transformed these strains with a non-fluorescent variant of GFP and ensured that each had a similar growth rate and expansion speed to the same strains with fluorescent GFP (Extended Data Fig. 3a–f). Equal mixtures of various strain pairs were inoculated at the centres of agar plates and their spatiotemporal abundance patterns were characterized (see Methods). Figure 2a shows the outcome of competition between mutant D and the ancestor 12 h after inoculation. Notably, the two strains dominated different spatial regions: the ancestor (pseudocoloured purple) dominated the interior whereas mutant D (pseudocoloured cyan) dominated the exterior. Repeating the competition process between mutant B and the ancestor (Fig. 2b), we found the opposite, with strain B dominating the interior and the ancestor dominating the exterior. The ratio of the calibrated fluorescence intensities, shown as the coloured solid lines in Fig. 2c, d, agree well with the ratio \( W \) of the densities of evolved cells over the ancestor (circles in Fig. 2c, d) as obtained from cell counting (Extended Data Fig. 4f), a direct measure of the relative fitness of the mutant at each location. This relative fitness profile was stable through much of the 24-h course of competition (Extended Data Fig. 4g–j).

As strain B expanded slower than the ancestor whereas strain D expanded faster (Fig. 1c), the competition results suggest a trend in which the slower strain dominates the interior and the faster strain dominates the exterior. This is in stark contrast to the ratio of cell densities from single-strain expansion dynamics (coloured dashed lines in Fig. 2c, d, derived from Extended Data Fig. 4b), which shows an advantage for the faster strain everywhere. Thus, the faster strain became disadvantaged in the interior only when grown in the presence of the slower strain, manifesting the ‘game-like’ nature of the underlying evolutionary process that, in this case, the fitness of a strain at a location depends on the presence and motility of competing strains.

Repeating the competition assay between the ancestor and each of the six mutants, we found that the slower strain dominates inside and the faster strain outside in each case (Extended Data Fig. 4k–m). The competition results can be concisely summarized by defining a crossover distance \( d \), at which the ancestor and the mutant have the same fitness: \( W(d) = 0 \). This is indicated as the vertical dashed line in Fig. 2c, d and Extended Data Fig. 4m. This crossover distance is plotted against the expansion speed of the corresponding mutant in Fig. 2e, with various regions shaded according to strain dominance: the faster strain for \( X > d \), and the slower strain for \( X < d \). To see whether the competition results represented by this ‘phase diagram’ are specific to the evolved strains, we repeated these studies using two synthetic strains, WL1 and WL2 (Supplementary Table 2), which allowed us to titrate swimming speed and expansion speed by using specific inducers without affecting cell growth (Extended Data Fig. 3g–l). When we competed the fluorescent versions of these strains with each other, we obtained a phase diagram (Extended Data Fig. 3m) that was very similar to that between the ancestor and evolved strains (Fig. 2e). This indicates that the latter represents a generic outcome of competition between strains with different motility characteristics, regardless of how these characteristics are changed.

**Modelling competitive expansion dynamics**

To gain more insight into the competition dynamics, we turned to a mathematical model of bacterial population expansion that includes the effect of cell growth along with the random and directed components of cell motion, based on well-characterized molecular interactions (Extended Data Fig. 5a). This model provides a quantitatively accurate description of the expansion dynamics for a single bacterial strain in soft agar (Extended Data Fig. 5b, c). We extended this model to describe competition between two strains that are assumed to respond to the same chemoattractant and grow at the same rate (Extended Data Fig. 6a), with different expansion speeds modelled by different parameters characterizing chemotaxis (see Supplementary Model for details and Supplementary Video 2 for an example of the dynamics). This model captured the spatial dominance pattern of the slow and fast strains observed after a long time (Extended Data Fig. 6b), as well as the time dependence of the crossover distance \( d \) (Extended Data Fig. 6c). Factors that favour the dominance of slower strains at smaller...
distances are attributed to shifts in balance among cell growth, forward movement, and back-propagation of pioneering cells from the population front. Using this competitive expansion model, we systematically computed the outcome of competition between an equal initial mixture of two strains, varying the expansion speed of one strain (the mutant, with speed $u'$) while holding that of the other (ancestor strain, with speed $u$) at a fixed value. Figure 2F shows the crossover distances $d(u, u')$ and the resulting phase diagram obtained for these competitions, which are similar to those observed experimentally (Fig. 2E, Extended Data Fig. 3M).

We also used the competitive expansion model to probe the dependence of spatial dominance for three strains. Consider three strains (a, b, c) with single-strain expansion speeds $u_a, u_b, u_c$ such that $u_a < u_b < u_c$. Let us find the region of dominance by strain b during three-strain competition. From the two-strain crossover distances $d_{ab} = d(u_a, u_b)$ and $d_{bc} = d(u_b, u_c)$ between strains a-b and b-c, respectively, the illustration in Fig. 3A clearly suggests that strain b will dominate in the region $d_{ab} < d < d_{bc}$. This is verified experimentally by directly competing three strains with different expansion speeds (that is, strains A, C, E in Fig. 3B). Thus the outcome of three-strain competition can be correctly predicted from the results of two pairwise two-strain competitions (in particular, the form of the crossover distance). We next show that this simple result can be generalized to predict the outcomes of the evolution experiments shown in Fig. 1.

An evolutionarily stable strategy

To connect to these evolution experiments, which involve potentially many strains with a continuum of expansion speeds, let us first consider the theoretical limit $u_a \to u_a$ and $u_b \to u_b$ (black arrows in Fig. 3A). In this case, the region where strain b dominates will be pinched and distributed narrowly around a special distance, $d_b^* = \lim_{u_a \to u_a, u' \to u_b} d(u_a, u'_b)$, the distance at which the strain with speed $u_b$ is dominant over other strains even if their speeds are infinitesimally different. As there is nothing special about strain b and its speed, this consideration suggests a much more general result: that for a strain with a single-strain expansion speed $u$, there is a special distance

$$d^*(u) = \lim_{u' \to u} d(u, u')$$

(1)

at which no other strain with a different speed can dominate.

Given the form of the crossover distance shown in Fig. 4A, the diagonal $d^*(u)$ (pink dashed line) as defined by equation (1) turns out to depend linearly on $u$ as shown in Fig. 4B. This simple result is reinforced by a more detailed mathematical analysis (Supplementary Analysis 1, 2), which further predicts that the slope of $d^*$ versus $u$ will be inversely proportional to the growth rate $\lambda$:

$$d^*(u) = u/\lambda$$

(2)

This form is confirmed by numerical simulation of the competitive expansion model performed at different growth rates (Extended Data Fig. 6D, E).

So far, equations (1) and (2) refer to the dominance of a strain in a 50:50 initial mixture of it with a competing strain. To apply these results on multi-strain competition to evolutionary dynamics, in which mutants may be generated at very low frequencies, it is necessary to recalculate the crossover distance $d(u, u')$ for a low frequency of the competing
strain. However, as we show in Extended Data Fig. 8, for two strains with comparable expansion speeds, their crossover distance is independent of the frequency of the competitor. Thus, equation (2) can be applied to competitions involving strains of arbitrarily small frequencies, including spontaneously generated mutants. Therefore, equations (1) and (2) describe an evolutionary stability criterion, that a strain with expansion speed $u$ is stable against invasion by mutants with different expansion speeds at position $d^*(u)$ as given by equation (2). We therefore refer to $d^*$ as the stability distance and $d^*(u)$ as the stability line.

The actual selection experiments performed (Fig. 1) pose a slightly different question from the evolutionary stability criterion just described: at a given selection distance $X$, what speed $u^*(X)$ is most fit? The answer is just the mathematical inversion of equation (2):

$$u^*(X) = \frac{X}{\lambda}$$

(3)

This result can be appreciated by examining a slice of the crossover landscape of Fig. 4a, for $(u, u') = X$ as shown in Fig. 4c. The stable speed selected is at the intersection of $d(u, u') = X$ (cyan line) and the diagonal $(u = u'$, pink line), indicated by the circle, as a strain with speed that deviates from the intersection (black arrows) is selected against; see legend for details. In the plot of the stability line (Fig. 4b), we added the teal-coloured secondary axes: at a given distance $X$, the selected speed $u^*(X)$ is obtained by following the teal arrows, whereas for a strain with a given speed $u$, its stability distance is obtained by following the black arrows.

![Image](https://example.com/image.png)

**Fig. 4** Stability line of the competition dynamics. Competition results for pairwise combinations of expansion speeds $(u, u')$ following the growth-expansion model (Extended Data Figs. 5, 6). a, The crossover distances $d(u, u')$ at which both strains are equally abundant (green surface). The considerations in Fig. 3a indicate that the diagonal (pink dashed line where $u = u'$) gives the stability distance $d^*$ of a strain with ancestral speed $u$. b, The relation $d(u')$ is linear and is called the stability line. This line has an orthogonal interpretation: following the teal arrows, it gives the expansion speed $u'$ that would be selected at position $X$. To verify this orthogonal view, we note that at a given distance, for example, $X = 15\text{ mm}$, there is a set of expansion speed combinations for which the crossover distances correspond to this distance $(d = X$, cyan line in a). A distinct speed $u'$ among this set is indicated in white, corresponding to the limiting value of $d$, when the expansion speeds of the two competing strains approach each other, defined mathematically from $X = \lim_{u \to u'} u^{-1} d(u, u')$. c, A strain with this special expansion speed $u^*(X)$ is stable against mutants with different speeds at distance $X$, according to the strain dominance pattern shown: different regions in this panel are assigned in the same way as in Fig. 2f. For $u > u'$, the ‘ancestor strain’ dominates where the green surface in a is below $X$, and the ‘mutant’ dominates where the green surface is above $X$; vice versa for $u' > u$. The expansion speed $u^*(X)$ is located where the phase boundary (cyan line) intersects the diagonal. Here, if the speed of one strain is increased or decreased (arrows), then it is selected against as the other strain dominates.

**Validation of the stability criterion**

To test the predicted stability line (equations (2), (3)), we designed two additional sets of evolution experiments for growth conditions that provided altered ancestral expansion speed and growth rate. First, we repeated the evolution protocol shown in Fig. 1a using the same growth medium (TB) but different agar densities. This changed the effective cell diffusion constant, resulting in a range of expansion speeds (purple squares, Fig. 5a) without affecting the growth rate. The evolution results obtained for five agar densities, each for five selection distances and different replicates, were highly reproducible (Extended Data Fig. 9a). From the evolved expansion speeds, we determined the stable selection distance for each agar density (Extended Data Fig. 9b, c). The stability distances obtained exhibited a linear relation with the ancestral expansion speeds, as predicted (Fig. 5b).

Next we plotted the expansion speeds obtained from different cycles of evolution (the data in Extended Data Fig. 9a) with the stability line in Fig. 5c–e for data from three different agar concentrations. Interpreting the stability line as the stable expansion speed $u^*(X)$ at the corresponding selection distance $X$ (Fig. 4b), the data from each evolution series (symbols of the same colour) are seen to approach the predicted final stable values.

Then we repeated all of the evolution experiments yet again, at various selection distances and agar densities, in a medium that supports approximately 50% slower growth (CAA; brown squares in Extended Data Fig. 9d). The expansion speeds of the evolved strains (Extended Data Fig. 9e) exhibited a similar pattern of changes to those obtained in TB (Extended Data Fig. 9a). The stability distances obtained for different ancestral expansion speeds (Fig. 5f) again followed a linear relation as predicted, with an approximately 70% increase in the slope, consistent with the dependence on cell growth rate given in equation (2).

To probe the generality of a stable expansion speed and its linear size-dependence, we investigated another mode of selection in silico using the multi-strain generalization of the competitive expansion model (Supplementary Analysis S5). In this mode of selection, a fraction...
null
Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Media and growth conditions

The TB medium contains 10 g tryptone and 5 g NaCl per litre. The M9 supplemented medium is based on the Knight laboratory’s recipe: 1× M9 salts, 0.2% casamino acids, 2 mM MgSO4, 0.1 mM CaCl2, and the carbon source is 0.4% (v/v) glycerol. M9 salts were prepared to be 5× M9 salts stock solution (in 11): Na2HPO4·30 g, KH2PO4·15 g, NH4Cl·50 g, NaCl·2.5 g. The Lucia-Bertani (LB) medium used in this study contains 2.5 g yeast extract, 5 g bacto tryptone, and 5 g NaCl per litre. For all the expansion experiments, the bacto-agar (BD, 214010) was added to the growth medium and the agar concentration varied from 0.2% to 0.3% (w/v).

To prepare semi-solid agar, the above growth medium was buffered to pH 8.0 with 0.1 M HEPES (pH 8.0), the pH variation was less than 0.3. Then, 10 ml of the above medium supplemented with different agar concentrations was poured into a 90-mm Petri dish, and allowed to harden at room temperature for 90 min. Unless otherwise stated, all other reagents were from Sigma. All experiments were carried out at 37°C. Plasmids were maintained with 100 µg/ml ampicillin, 25 µg/ml kanamycin, 25 µg/ml chloramphenicol, or 50 µg/ml spectinomycin.

Strains and plasmid construction

The ancestor E. coli CLM strain used in this study was provided by A. Danchin (AMABiotics, France). All strains used in this study are listed in Supplementary Table 2. Strains are available upon request. Oligonucleotides used are listed in Supplementary Table 4.

The cheZ-titratable strain W1L1 was constructed as described previously. In brief, the bla::Ptet-tetR-cheZ at attB site was amplified from the pMD19-T-Amp-R-T::Ptet-tetR-cheZ with primers PR29 and PR30 and inserted into the CI1 (ΔcheZ, Δlac) chromosomal attB site by recombineering with the aid of pSIM5. To construct the cheZ-titratable strain W1L2 (ΔcheZ, Δlac, blalacI-tetR-cheZ at attB site) the plasmid pMD19-T-Amp-R-T::Ptet-tetR-cheZ was first constructed. In brief, the plasmid was constructed by inserting PCR-amplified P::lacIwith Xhol and HindIII restriction sites from Q5 into the corresponding restriction sites of the pMD19-T-Amp-R-T::Ptet-tetR-cheZ, finally the Ptet-tetR was replaced with the P::lacI. Then the blalacI-tetR-cheZ feedback loop was amplified from the plasmid and inserted into the attB site by recombineering.

The derived clpx (G371S, *425Q, P67S), the 66th amino acid Y was mutated into C by overlapping PCR. In brief, the targeted locus was first replaced with blax by λ-Red, the evolved alleles were then introduced by CRISPR-Cas9 targeting the blax gene. Procedures used for the allelic exchange are as follows. First, the pCASsac plasmid containing the cas9, λ-Red and sacB was electroporated into the ancestor and strains with the kanamycin resistance were selected. Then, PCR products containing the -50 bp sequence homologous to the targeted locus on each side of the T1-AmpR::Ptet-blax cassette were amplified from pMD19-T-Amp-R-T::Ptet-tetR plasmid. The purified DNA fragments were then electroporated into the ancestor strain with pCASsac. The T1-AmpR::Ptet-blax cassette was integrated into the targeted locus with the aid of the λ-Red from pCASsac and the ampicillin-resistant clones were selected. These cells were identified by colony PCR. Subsequently, the pTargetF-AmpR plasmid carrying the N20 (20-bp region complementary to the target region) that was targeted to the bla region was obtained by inverse PCR with the primers PCC35 and PCC36 from the pTargetF followed by self-ligation. PCR products containing the evolved alleles were amplified from the evolved strains with appropriate primers listed in Supplementary Table 4. The pTargetF-AmpR plasmid carrying the sgRNA, lacIq-P::rec promoter guiding the PMB1 replication of pTarget and the PCR fragments with the evolved allele were co-electroporated into the ancestor with T1-AmpR::Ptet-blax cassette at the targeted locus and pCASsac plasmid.

Cells grown on LB agar containing kanamycin and spectinomycin were identified by colony PCR and DNA sequencing. Finally, the constructed strains both contained the pTargetF-AmpR and the pCASsac plasmids. The pTarget-AmpR was first cured by a second round of genome editing. Cells grown on LB agar with kanamycin that did not grow on LB agar with spectinomycin were selected. These selected cells were picked into 2 ml of LB medium with 5g/l glucose for overnight culture; cells grown on LB agar containing 5g/l glucose and 10g/l sucrose were selected as the final strains.

The fluorescence plasmid PZA31-Ptet-M2-GFP was from the Hwa laboratory. To construct a loss-of-function non-fluorescent GFP mutant NFP1, the 66th amino acid Y was mutated into C by overlapping PCR. In brief, PZA31-Ptet-M2-GFP was reverse amplified with a pair of complementary primers GFP-Y66C-f and GFP-Y66C-r. The PCR product was purified and treated with DpnI, gel purified, ligated, and then transformed into the DH5α-competent cells. The plasmid was verified by sequencing.

Evolution experiment procedures

First, the ancestor strain from the ~80°C stock was streaked onto an agar plate and cultured at 37°C overnight. Three to five single colonies were picked into 2 ml of the corresponding growth medium and cultured at 37°C overnight. Second, the overnight culture was diluted into 2 ml pre-warmed fresh growth medium with a ratio of 1:100 the next morning. Bacteria were then cultured to the mid-log phase (OD600 was around 0.2–0.3), and 2 ml of the ancestor strain was inoculated at the centre of a semi-solid agar plate and incubated at 37°C for 24 h. Cells grew and migrated, occupying the whole semi-solid agar plate (marked as cycle 0). We picked 2 ml of the agar–cell mixture from site A, B, C, or E of this master plate (with a radius of 5, 10, 15, 20, or 25 mm away from the inoculum, respectively), directly inoculated onto fresh semi-solid agar plates (marked as A, B, C, D, or E series, correspondingly), and incubated at 37°C for another 24 h, marked as cycle 1. Then 2 ml of the agar–cell mixture was picked at site A from plate A, site B from plate B, site C from plate C, site D from plate D, and site E from plate E, inoculated onto the centre of the fresh semi-solid agar plates, and incubated at 37°C for 24 h, marked as cycle 2. This process was repeated for 50 cycles, with selection always kept at the same radius. Bacteria from site A, A, B, C, D, or F (with a radius of 3, 5, 10, 15, 20, or 35 mm away from the inoculum, respectively) were passaged for 40–50 cycles following the above process in M9 minimal medium supplemented with glycerol and casamino acid medium (M9 + glycerol + CAA). For the evolution experiment carried out in semi-solid LB agar plates, 2 μl of cell–agar mixture picked from site A, D, or F (with a radius of 5, 20, or 35 mm away from the centre of the semi-solid agar plate, respectively) were transferred onto fresh semi-solid agar plates. Considering the fast growth rate and expansion speed in this medium, the cycle of the culture was shortened to 12 h. For the evolution experiment carried out in M9 minimal medium supplemented with glycerol, A’, D, or F (with the radius of 3, 20, 35 mm away from the centre of the semi-solid agar plate, respectively) were transferred every 72 h, and this process was repeated for 30 cycles. The samples of the evolving populations were stored at ~80°C right after well mixing the agar–cell mixture from indicated sites with an equal volume of 40% (v/v) glycerol. All the evolution experiments were performed in at least three replicates. The >5 × 103-fold daily growth corresponds to ~12.3(log10[5 × 103]) generations of doubling. The number of doublings during each cycle was estimated as follows: cell density was about 3.54×107/ml in TB medium and 1.93×109/ml in M9 + glycerol + CAA after migrating in semi-solid agar for 24 h. The initial inoculum cell density was about 7.43×105 per ml for TB and 4.94×108 per ml for M9 + glycerol + CAA. These numbers were counted using flow cytometry (see below).
Expansion speed measurement
The semi-solid agar plate was illuminated from below by a circular white LED array as described previously\(^1\) and imaged at 30 min or 1 h intervals using a Canon EOS 600D digital camera. Images were analysed using ImageJ and a custom-written image analysis script using MATLAB. A circle was fitted to the intensity maximum in each image and the area (A) of the fitted circle was determined. The radius (r) of the colony was calculated as \(r = \sqrt{A/\pi}\). The maximum expansion speed was calculated using a linear fit over a sliding window of at least four time points, with the requirement that the fit has an \(r^2\) greater than 0.99.

Growth rate measurement
Growth rates of the evolved strains were measured in a 100-ml flask with 20 ml corresponding growth medium at 37°C, 150 rpm. The procedure was as follows. First, the isolated bacteria from −80°C stock was streaked onto the agar plate, and cultured at 37°C for 4 h. Second, 3–5 single colonies were picked and inoculated into 2 ml corresponding growth medium and cultured overnight. The overnight culture was diluted into 2 ml pre-warmed medium with a ratio of 1:100 the next morning and cultured to log phase. The log phase culture was successively diluted into 20 ml pre-warmed growth medium, the final OD\(_{600}\) was about 0.02–0.05. OD\(_{600}\) was measured using a spectrophotometer reader every 12 min (for TB) or 15 min (for M9). At least three doubling times were recorded. Maximum growth rates were calculated using an exponential fit over a sliding window of at least five time points, with the requirement that the fit has an \(r^2\) greater than 0.99.

Competition assay
Head-to-head chemotactic competition was observed using a pair of strains with competition either between the ancestor strain and one evolved strain, or between the two cheZ titratable strains (Extended Data Fig. 3) induced with aTc or IPTG. To allow observation by fluorescence microscopy, strains carrying either GFP- or NFP-expressing plasmids were prepared (Extended Data Fig. 3). Each competition was repeated for both combinations of plasmids (for example in the competition run ancestor versus A\(_c\), the growth and expansion of the ancestor strain was observed for the run with Anc\(_c\) versus A\(_c\) as well as for the run Anc\(_c\) versus A\(_c\)). Both the fluorescence intensity and the cell number of each pair of competing strains at different positions and different time points were measured to characterize the competitiveness of the cells in semi-solid agar. The competition experiments for ancestor versus an evolved strain were initiated as follows: three to five single colonies of the isolated evolved strain and the ancestor with the GFP/NFP were cultured to log phase (OD\(_{600}\) was around 0.20) separately. Two types of the combined mixed strains were prepared: the evolved strain with GFP was mixed with the Anc\(_c\) while the evolved strain with NFP was mixed with the Anc\(_c\) in a 1:1 ratio. Next, 2 μl of the two types of the combined mixture were inoculated onto the centre of pre-prepared semi-solid agar plates separately and allowed to expand at 37°C for 30 min or 1 h intervals using a Canon EOS 600D digital camera. Images were collected from the agar plate, and cultured at 37°C overnight. Second, 3–5 single colonies were picked and inoculated into 2 ml corresponding growth medium and cultured to log phase. The log phase culture was successively diluted into 20 ml pre-warmed growth medium, the final OD\(_{600}\) was about 0.02–0.05. OD\(_{600}\) was measured using a spectrophotometer reader every 12 min (for TB) or 15 min (for M9). At least three doubling times were recorded. Maximum growth rates were calculated using an exponential fit over a sliding window of at least five time points, with the requirement that the fit has an \(r^2\) greater than 0.99.

Fluorescence intensity as a function of cell density was calibrated as follows. The Anc\(_c\), A\(_c\), B\(_c\), B\(_c\), C\(_c\), D\(_c\), and E\(_c\) strains were cultured in TB medium to mid-log phase: OD\(_{600}\) was around 0.20 for each strain. A total of 200 μl culture was collected and concentrated to 1.6 × 10^6 cells/ml in the TB medium with 2 mg/ml kanamycin for each strain. Then, a serial dilution of the above concentrated sample into TB medium with 2 mg/ml kanamycin was carried out. Subsequently, the diluted samples were mixed with 0.277% (w/v) TB agar containing 2 mg/ml kanamycin in a ratio of 1:9; 10 ml of the cell–agar mixture was poured into a 9-cm Petri dish and allowed to solidify at room temperature for 90 min, and 100 μl of the cell–agar mixture was used for cell counting with a flow cytometer. The fluorescence intensity of the above cell–agar mixture plate was measured using a fluorescence microscope, as for the two-strain competition assay. Then the relationship between the fluorescence intensity of the cell in semi-solid agar and the cell density was plotted.

The initial and final ratios of the two competitors were measured by cell counting with a flow cytometer (Beckman, Cyto-FLEX). In brief, samples were first fixed with pre-cooled cell counting buffer (0.9% NaCl with 0.12% formaldehyde). Subsequently, the fixed samples were diluted as necessary with staining buffer (cell counting buffer with 0.1 μg/ml DAPI) before the flow cytometer analysis. Finally, the stained samples were counted with the flow cytometer. The flow rate was 30 μl/min and at least 50,000 cells were collected. The DAPI-stained particles were deemed to be the bacterial cells, and the DAPI-positive cells were separated into two groups (GFP and NFP) through the FITC channel. The fitness \(W_{i}\) of strain \(i\) (relative to the ancestor) is defined as the ratio of density at distance \(d\) (and a sufficiently long time \(t\)) over the initial inoculent density, \(p_i(d,t)/p_i(0,0)\), relative to the same ratio for the ancestor: \(W_i(d) = [p_i(d,t)/p_i(0,0)]/[p_m(d,t)/p_m(0,0)]\).

Quantitative real-time RT–PCR
A volume of 1 μl of the log phase bacteria (OD\(_{600}\) ~0.2) from each condition was immediately mixed with 2 ml RNA protect Bacteria Reagent (Qiagen). Total RNA was extracted using the RNeasy Mini kit (Qiagen) according to the manufacturer’s protocol. The RNA yield and purity were checked using a NanoDrop 2000c spectrophotometer (Thermo Scientific), and the absence of genomic DNA contamination was confirmed by PCR. About 500 ng RNA was reverse transcribed, using a PrimeScript RT reagent kit with gDNA Eraser (Takara) according to the manufacturer’s protocol. Reactions without reverse transcriptase were conducted as controls for the following qPCR reactions. Then the cDNA samples were diluted 1:25 with PCR-grade water and stored at −20°C until use. SYBR Premix Ex Taq (TliRNaseH plus) (Takara) was used for qPCR amplification of the cDNA. Then, 5 μl diluted cDNA sample, 200 nM forward and reverse qPCR primers, 10 μl SYBR Premix Ex Taq, and up to 20 μl PCR-grade water were mixed in a well of a 96-well PCR plate (BIO-RAD). The non-template control (NTC), containing sterile water instead of cDNA template, was included during each qPCR experiment to check the purity of the reagents. Each reaction was performed in triplicate. The qPCR reactions were performed using a Bio-rad CFX connect Real-Time system with the following program: 30 s at 95°C and 40 cycles of denaturation (5 s at 95°C), annealing, and elongation (30 s at 60°C). Data were acquired at the end of the elongation step. A melting curve was run at the end of the 40 cycles to check the specificities of the accumulated products. To calculate PCR efficiency, standard curves were made for the target gene by using serially diluted cDNA samples as the templates. 16S rRNA was used as the reference gene to normalize the expression level.

Single-cell tracking
Single-cell tracking was performed as described previously\(^9\) with minor modifications. A custom MATLAB script was used to control the automated stage of the microscope (Nikon Ti-E) via the Micromanage interface\(^8\). Movies were sequentially acquired using an Andor Zyla 4.2 sCMOS camera at 10 frames/s through a 10× phase-contrast objective.
Whole-genome sequencing and analysis

Whole-genome sequencing was performed using the Illumina platform, obtaining an average >100× coverage. Samples directly collected from the TB evolution experiment in 0.25% agar were used for sequencing. In brief, the isolated clones were grown in the growth medium as described previously and harvested at stationary phase. For the population samples, 100 μl of the frozen cell–agar mixture from the 50th cycle was grown in 3 ml growth medium and harvested at the 10th hour. Cellular DNA was extracted using a genomic DNA purification kit (Tiangen) according to the manufacturer’s protocol. Whole-genome libraries were prepared and sequenced on the Illumina HiSeq X10 by BGI. The genome sequence of E. coli MG1655 (NCBI: NC_000913.3) was used as the reference sequence. All sequencing data were analysed using the BREFSEQ pipeline\(^\text{35}\) supported by Python. A subset of the identified mutations was resequenced using Sanger sequencing for confirmation.

Numerical simulations of competition dynamics and evolution

The growth–expansion model\(^\text{41}\) (Extended Data Fig. 5) was extended to analyse competition and evolutionary dynamics during expansion, see Extended Data Fig. 6 for introduction and defining partial differential equations (PDE), and Supplementary Model for mathematical details. Numerical solution of the partial differential equations was done using an implicit scheme using Python 2.7 and the PDE solver module FiPy\(^\text{43}\). Integration over time was typically performed with time steps \(dt = 0.25\) s, and a grid resolution with spacing \(dx = 10\) μm. Simulations were performed using a custom-made Python code, which is available via GitHub at https://github.com/jonascremer/chemotaxis_simulation. Used parameter sets are provided in Supplementary File simulationparameters.txt, see Supplementary Model 5 for usage.

Data availability

Sequencing data have been deposited to the NCBI Sequence Read Archive (SRA), accession PRJNA559221. Other major experimental data supporting the findings of this study are available within the paper and Supplementary Information. Simulation data can be generated with the custom-made code and the parameter sets provided.

Code availability

Custom-made simulation code is available via GitHub at https://github.com/jonascremer/chemotaxis_simulation.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.
Extended Data Fig. 1 | Characteristics of cell motility and expansion speed for the ancestor and for mixed populations of evolved cells. a, Navigated range expansion by bacteria on a soft agar plate involves a group of pioneering cells (cyan) at the population front that move outwards towards uncolonized space (grey) in response to a signalling cue. During the outward migration, they replicate and leave offspring behind the front. The offspring do not move outwards, but settle wherever they are deposited and grow exponentially (green) until they reach carrying capacity (brown).

b, Density profile of a population of ancestor cells containing GFP (Anc G) at various times after inoculation at the centre of a semi-solid TB plate with 0.25% agar incubated at 37 °C. The population expanded at a defined expansion speed, covering positions A–E well before selection at 24 h. The density profiles were obtained using confocal microscopy as described previously. Cellular characteristics were not significantly affected by the expression of GFP (Extended Data Fig. 3).

c, The experiment described in Fig. 1a was carried out independently for five distinct selection positions, at distances X_A, X_B, X_C, X_D and X_E, ranging from 5 to 25 mm from the centre. Three independent lineages of this experiment were propagated in parallel (Extended Data Fig. 2a). Selected samples at cycle n of series S in lineage l are referred to as Sl_n in Supplementary Table 2.

d, Front position versus time for the ancestral strain CLM (purple) and populations of evolved cells obtained at the 50th cycle of each of the five evolution series from lineage 1. For each mixed population, collected cells were grown to mid-exponential phase (OD_{600} = 0.20), and 2 μl of the batch culture was inoculated at the centre of the same TB plate as in b and incubated in the same way. The agar plates were photographed at different times after inoculation and the radius of each expanding population was deduced from the area measured using ImageJ. The expansion speed of each population (Fig. 1b) was obtained as the slope of the linear fit of the data after t = 3 h.

e, Growth rates of population samples from each of the five selection series from lineage 1 at various evolution cycles. f, Growth rates of population samples from each of the three selection series (lineage 1) at various cycles of evolution experiments in CAA medium. g, Scatter plot of growth rates and expansion speeds for single strains isolated from frozen samples taken at various cycles of CAA evolution experiments. The evolution cycles are indicated by circles. h, i, Mean run speed (h) and tumble bias (i) of population samples from each of the three selection series from lineage 1 of evolution experiments carried out in TB medium at various evolution cycles. At least 10,000 cells were subjected to single-cell tracking analysis for each experiment (see Methods). Error bars represent s.d. j, Expansion speeds of several strains, each carrying an identified mutation, plotted against the corresponding evolved strains from which the mutations were identified. The mutations are indicated in the legend (see Supplementary Table 1). mutH does not affect motility and serves as a control. Experiments were repeated independently three times for b, d and twice for e, f with similar results. In h, i, data are mean ± s.d. for a single biological replicate, n = 10,000 cells analysed. In j, data are means for n = 3 biological replicates (s.d. error bars are smaller than the symbols).
Extended Data Fig. 2 | Experimental evolution of expansion speed in different growth media. a–d. Expansion speeds of population samples from each selection series are shown at various cycles of evolution experiments carried out in TB (a), LB (b), CAA (c) and glycerol-containing minimal medium (d). The agar density was 0.25% (w/v). All three lineages of each medium showed similar trajectories. The absence of a chemoattractant in the glycerol minimal medium leads to a very different outcome compared to the experiments carried out in complex media (a–c). In the absence of a chemoattractant, cells follow simple Fisher–Kolmogorov dynamics. In line with the absence of a growing population trailing the front, no decrease in swimming behaviour was observed over time. However, as previously investigated, slower swimming behaviour might be selected for when density-dependent growth effects or strong tradeoffs between swimming and cell growth exist. Experiments in a, b, d were carried out once with three biologically independent repeats with similar results; experiment in b was repeated independently twice with similar results.
Extended Data Fig. 3 | Fitness effects for cells expressing GFP and its non-fluorescent variant NFP. a, b, Growth rates (a) and expansion speeds (b) of the ancestor and six mutant clones harbouring GFP, NFP, or no plasmid, along with the corresponding mixed populations from which the six mutant clones were isolated. Plasmids harbouring constitutive GFP or NFP expression (PZA31-Ptet-M2-GFP or PZA31-Ptet-M2-NFP, respectively) were transformed into the six evolved strains A, B, B’, C, D, and E, to form AG, AN and so on (Supplementary Table 2). c, Fluorescence intensity as a function of cell density. The cell growth of cells expressing GFP was arrested by adding 2 mg ml\(^{-1}\) kanamycin, and cells were concentrated to 1.6 \times 10^{10} cells per ml. Subsequently, serial dilutions were carried out. For each cell density, cells were vigorously mixed with pre-warmed 0.277% (w/v) TB agar containing 2 mg ml\(^{-1}\) kanamycin at a ratio of 1:9 and poured into three Petri dishes with 10 ml each. All dishes were allowed to harden at room temperature for 90 min. The cell–agar mixture was subjected to cell counting by fluorescence-activated cell sorting (FACS) and the fluorescence intensity of the agar plate was detected using a fluorescence microscope. d, The three-dimensional structure of GFP with the predicted position of the loss-of-function mutation introduced, Y66C. The shown crystal structure is based on the GFP protein sequence aligned to PDB ID: 1GFL.\(^{47}\) The rectangle shows the mutation position. e, Ancestor cells harbouring constitutive GFP or NFP expression (Anc and Anc, cells, respectively) as viewed by fluorescence microscopy; the images verify the loss-of-function mutation of GFP. f, The relative fitness W of Anc and Anc, cells at different distances from the agar plate centre. Cells harbouring the GFP or NFP plasmid were equally mixed and inoculated at the centre of 0.25% TB plates. Cell–agar mixtures picked at various distances were subjected to cell counting by FACS 24 h after inoculation and the ratio was reported as W (see Methods). g, h, The genetic circuit and characteristics of the cheZ-titratable strains WL1 (g) and WL2 (h). The expression of cheZ depends on the control of a Ptet-tetR (WL1) or Plac-lacI (WL2) feedback loop and the native cheZ was seamlessly removed (see Methods). Relative cheZ mRNA levels change around twofold under different concentrations of inducers (\(\alpha\text{Tc for WL1 and IPTG for WL2}\). Relative cheZ expression levels, expansion speed, and growth rates of WL1 (g) and WL2 (h) under various concentrations of the respective inducers are shown next to the circuits. Horizontal dashed lines show the corresponding values for the ancestor strain. i–l, Growth rates (i, k) and expansion speeds (j, l) of the \(\alpha\text{Tc}\)-titratable cheZ strain WL1 (i, j) and its derivatives expressing GFP (WL7) or NFP (WL8), and the IPTG-titratable cheZ strain WL2 (k, l) and its derivatives expressing GFP (WL9) or NFP (WL10). m, Circles indicate crossover distances between fluorescent derivatives of two cheZ-titratable strains (WL1 and WL2), with WL1 strains induced at a fixed concentration of its inducer (0.5 ng ml\(^{-1}\) \(\alpha\text{Tc}\)), and WL2 strains induced at various IPTG concentrations (expansion speeds (h, l) shown on the x-axis). The background colours again indicate dominance by WL1 (purple) or WL2 (green). In a, b, g–l, data are mean ± s.d. for \(n = 3\) biologically independent repeats (individual data points shown as circles). Error bars in g, h were smaller than the symbols. Experiments shown in c, e, f, m were repeated independently three times with similar results.
Extended Data Fig. 4 | Results of two-strain competitions. **a**, Expansion speeds and growth rates of ancestor, strain B, and strain D (Supplementary Table 2). **b**, Calibrated fluorescence intensity profiles of singly grown strains 12 h after inoculation at the centre of 0.25% TB agar plates. Unless noted otherwise, fluorescence intensity is normalized according to the calibration curve shown in Extended Data Fig. 3c; the relative value refers to $5 \times 10^7$ cells per ml. **c**, Relative fluorescence intensities obtained at various times for the fluorescent mutant strains BG (orange) and DG (cyan) and the ancestor AncG (purple), each grown singly on TB plates. The faster strain has higher fitness everywhere. **d**, Raw photographs (top) and fluorescence intensity profiles (bottom) before and after merging of a representative two-strain competition between the fluorescent derivatives of the ancestor and strain D 12 h after initial equal inoculation. We used plasmids GFP and NFP in this study to distinguish the two strains from each other in the head-to-head competition, as there is no systematic influence caused by the expression of GFP or NFP (Extended Data Fig. 3). Top (from left to right): competition between DG and AncN, competition between DN and AncG, and merged photograph in pseudocolor (DG, cyan; AncG, purple). Bottom: corresponding relative fluorescence intensity profiles. **e**, As in **d**, but with strain B instead of strain D. Evolved strains and the ancestor with GFP or NFP were cultured to log phase separately. Two types of the combined mixed strains were prepared and the evolved strain with GFP was mixed with AncG while the evolved strain with NFP was mixed with AncN in a 1:1 ratio. Subsequently, 2 μl of the two types of combined mixture was inoculated onto the centre of pre-warmed semi-solid agar plates separately and allowed to expand at 37 °C for up to 24 h. Photographs and fluorescent intensities of the evolved strains and ancestor with the GFP reporter from these two plates after the expansion were taken at various times and merged (see Methods). **f**, Relative fitness $W_B$ obtained as a ratio of the direct cell count of the fluorescent derivatives of the ancestor and strain B, inoculated at the centre of an agar plate. From left to right: competition between BG and AncN, competition between BN and AncG, and averaged curve of both (the relative fitness $W_B$). The competition experiments are the same as in e. The initial and final ratios of the two competitors were measured by cell counting with a flow cytometer (see Methods). **g**, i, The spatiotemporal development of the bacterial density profiles, indicated by fluorescence intensities, for the competition between the ancestor (purple) and strain D (cyan; g) or strain B (orange; i) at various times during the 24-h competition. Beyond the initial period (~6 h), the crossover distance could be clearly defined (vertical dashed line) and was practically time-independent, with the slower strain gaining advantage in the interior and the faster strain gaining advantage in the exterior. **h, j**, Relative fitness values $W_D$ (h) and $W_B$ (j) taken as the ratio of the fluorescence intensities (mutant:ancestor) at various distances for 6 h and beyond. The vertical dashed lines indicate the crossover distance $d$, where the densities of the two competing strains are equal ($W_i = 1$). Thus, the crossover distance was fixed shortly after the initial period. **k–m**, Fluorescence intensity profiles (k, l), and relative fitness $W$ (m) of representative two-strain competitions between the ancestor (black solid line) and evolved isolates. The data were taken 12 h (k, m) or 24 h (l) after co-inoculation of equal initial mixtures of the two competing strains at the centre of 0.25% TB agar plates, showing that the slower strain spatially outcompetes the faster strain within the crossover distance $d$ (dashed lines). See Supplementary Table 2 for strain information. Experiments in **a–l** were repeated independently three times with similar results. In **m**, the mean ± s.d. of $n = 3$ biologically independent repeats is shown.
Extended Data Fig. 5 | The growth–expansion dynamics of a single strain.

\( \frac{\partial x}{\partial t} = \text{diffusion} - \nabla \cdot (v \cdot \rho) + \lambda(n) \cdot \rho \)  
\( v = c \cdot \nabla \log \left( \frac{1 + a}{K_I} \right) \)  
\( \frac{\partial a}{\partial t} = D_0 A - \mu(a) \cdot \rho \)  
\( \frac{\partial n}{\partial t} = D_0 n - \frac{\lambda(n) \cdot \rho}{Y} \)

\( a \) is the concentration of an attractant that cells can sense and move towards; and \( n \) is the concentration of a nutrient that cells consume to grow. Following the spirit of the classical model introduced by Keller and Segel (the KS model), cells can move in a random, undirected manner (effective diffusion constant \( D_0 \) and diffusion term in equation (S1)) and along the gradient of the signalling molecule (the attractant gradient represented by the convection term in equations (S1) and (S2)), which is generated by cellular consumption (equation (S3)). To account for observed density profiles and their evolution over time, three additional aspects beyond the KS model are important to consider. First, cells can detect and respond to attractant gradients in a scale-free manner only within a limited range between \( K_I \) and \( K_A \), the lower and upper cutoffs describing the molecular limitations of attractant sensing as specified in equation (S2). \( D_0 \) denotes the molecular diffusion of the attractant and nutrient within the agar. The chemotactic parameter \( c \) denotes how cells translate the detected attractant gradient into directed movement. Second, cells grow throughout the expansion process, as described by the growth term in equation (S1), with growth rate \( \lambda \). Third, growth relies not on the presence of the attractant but on the presence of nutrients. We model the latter dependence by the nutrient field \( n(x,t) \) and a yield factor \( Y \) (equation (S4)). The distinct treatment of the roles of nutrient and attractant is designed to model the dynamics of bacterial cells in complex media, where non-chemotactic components in the complex media are designated as the nutrient, while the (minor) chemotactic components of the complex media are reflected by a low concentration of a single attractant; a detailed discussion and validation of the model, including comparison to other models of chemotactic migration, has been published previously.

\( a \), Emerging density profile (green solid line) and attractant concentration (magenta dashed line) in the GE model. At the front is a density bulge or peak, within which the attractant profile drops steeply, guiding cells to do chemotaxis. Directed movement of cells in the front bulge is coupled to an exponentially rising trailing region. In this region, cells grow and swim randomly, but there is no directed cell movement there owing to the low attractant concentration (\( a < K_I \)). Note that in cases in which multiple attractants are present in the medium, the population typically exhibits multiple rings, one corresponding to the exhaustion of each attractant. For these cases, the trailing region involving cell growth but no chemotaxis would correspond to the region inside the innermost ring, after the exhaustion of the last attractant. Thus, our model with a single attractant does not attempt to describe the movement of the outer rings, but models the transition of the density profile from the innermost ring to the exponential trailing region. As we describe in Supplementary Analyses 1, 2, the dynamics in this transition region determine strain dominance in multi-strain competition processes.

\( c \), Stable expansion of the population can be explained by a balance between growth of cells in the front bulge and leakage of cells out of the front due to random movement (cell diffusion). For illustration, consider the green dashed line indicating the density profile at a given earlier time. With only directed movement (along attractant gradient) and cell growth, the front peak would be higher at a later time, as illustrated by the brown dashed line. Instead, growth in the front bulge is compensated by back-diffusion of cells away from the front (green solid line indicating a later time). This compensation mechanism is further shown in the cylinder plots below, illustrating that the observed stable expansion dynamics (constant expansion speed, constant peak density) results from a dynamic balance between growth at the front and back-diffusion. The exponential density profile in the trailing region results from a combination of the steady outward movement of the front and a steady exponential growth of cells leaked out of the front.
Extended Data Fig. 6 | Model of two-strain competitive expansion dynamics.

**a**, The GE model of bacterial range expansion 14 (Extended Data Fig. 5) is extended to several strains of bacteria for which the densities are denoted by \( \rho_i(x,t) \), with \( i \in \{1,2\} \) for two strains. The different bacterial strains are assumed to consume the same nutrient \( n \) and grow at the same rate \( \lambda(n) \), in accordance with Monod’s law. They also sense and consume the same signalling molecule (the attractant \( a \)) at the same rate \( \mu(a) \). The random motion of each bacterial strain is described by a diffusion term, with effective diffusion coefficients \( D_i \) for strain \( i \) (equations (S5), (S6)). The spatial attractant profile \( a(x,t) \) (resulting from bacterial consumption, equation (S8)) leads to directed motion which is modelled by a drift term in equations (S5) and (S6). The dependence of the drift velocities \( v_i \) on the attractant profile is given by equation (S7), which describes a range of proportional sensing \( v_i \propto \partial_a a(x,t) / a(x,t) \) for \( K_i < a(x,t) < K_A \). The magnitude of the chemotactic response is parametrized by the chemotactic coefficient \( c_i \) for strain \( i \). Finally, the dynamics of the nutrient and attractant are described by equations (S8) and (S9), with \( D_0 \) characterizing molecular diffusion.

**b**, Outcomes of competitive expansion dynamics, showing the density profiles of two strains 24 h after inoculation with equal mixtures at the origin. Competition is run for one strain resembling the ancestor (black line, with expansion speed \( u_{anc} = 6 \text{ mm h}^{-1} \)) and the other strain resembling a mutant, with expansion speed \( u_{mut} \) increasing from left to right as indicated above the plots. The observed changes in crossover distances are minimal when growth-rate differences are minimal (for example, <5% as we observed experimentally for TB; Extended Data Fig. 4g) but become substantial when growth-rate differences become large (Supplementary Analysis 3).

**c**, The stability distance \( d^* \) as in **d** but for an alternative model formulation of chemotactic movement 50 with different functions describing sensing and the directed movement along gradients. See Supplementary Analysis 4 for more details. Despite the model changes, a linear relation with growth-rate-dependent slopes was still observed.
Extended Data Fig. 7 | Two-strain competition and crossover dynamics. The results in Fig. 2 show that competition between two strains with the same growth rate and different single-strain expansion speeds generally gives rise to a distinct spatial structure where the slower strain dominates inside and the faster strain dominates outside, separated by a crossover distance that depends on the two speeds. Here we explain how this feature arises from the underlying GE dynamics. a, Dynamics simulation using equations (55)–(59) for two strains differing only in their chemotactic coefficients \( c_{\text{slow}} = 652 \, \mu m^2 s^{-1} \) and \( c_{\text{fast}} = 727 \, \mu m^2 s^{-1} \), with corresponding single-strain expansion speeds \( u_{\text{slow}} = 4.95 \, mm h^{-1} \) and \( u_{\text{fast}} = 5.17 \, mm h^{-1} \); other parameters, equal for both strains, are provided in Supplementary Table 3. The density profile of the strain with faster single-strain expansion speed is green, the other purple. Red circles indicate crossover distances, which freeze at a fixed position in the laboratory frame (dashed red line) after the troughs of the two density bulges cross. This feature (see also Extended Data Fig. 4g,i) allows us to define a simple crossover distance \( d \) that is time-invariant. To understand the crossover dynamics in a, we first provide a qualitative explanation. While the two strains would individually expand at different speeds, the competition dynamics involves a single co-migrating population. This is because the two strains chase after the same attractant profile, which can recede only at a single speed \( u^* \). Because the front is moving faster than the speed at which the slower strain can stably propagate, the slower strain is gradually depleted from the front. As shown by the cylinders, at early times (before the front has reached the crossover distance and where the abundances of the two strains in the front bulge are comparable), the leakage flux of slower cells (purple) at the back exceeds that of the faster cells (green). As the two strains grow at the same rate behind the front, the slower strain dominates there. Because of its faster leakage, the slower strain is preferentially depleted from the front bulge. Subsequently, the leakage flux of the slower strain drops below that of the faster strain despite the faster leakage rate of the slower strain, owing to its reduced abundance. From there on, the back is dominated by the faster strain.

At the crossover distance, the two leakage fluxes are the same. b, A simple analysis captures key features of the crossover dynamics and leads to a time-invariant crossover distance \( d \), \( (u_{\text{slow}}, u_{\text{fast}}) \) for two strains with single-strain expansion speeds \( u_{\text{slow}} \) and \( u_{\text{fast}} \), and shows how this crossover distance leads to an evolutionarily stable selection distance \( d^* \) defined in the limit \( u_{\text{slow}} \to u_{\text{fast}} \) (Figs. 3, 4). The two density profiles shown in a, \( \rho_{\text{slow}} (x, t) \) and \( \rho_{\text{fast}} (x, t) \) in the ‘laboratory frame’ coordinate \( x \), are shown in b in the frame moving with speed \( u_{\text{fast}} \). In this moving frame where the spatial coordinate is \( y = x - u_{\text{fast}} t \), the density profile of the faster strain (green line) is approximately stationary: \( \rho_{\text{fast}} (x, t) = \rho_{\text{fast}}^* \) where \( \rho_{\text{fast}}^* \) is the stationary solution of the single-strain dynamics in the laboratory frame moving at speed \( u_{\text{fast}} \). Defining the position of the trough to be \( y = 0 \) in the moving frame, \( \rho_{\text{slow}} (y, t) \) has a bulge for \( y > 0 \) and a trailing exponential for \( y < 0 \). In this frame, the density profile of the slower strain is generally not stationary, and is written as \( \rho_{\text{slow}} (y, t) = \rho_{\text{slow}}^* (y - dt^*) \). As described in Supplementary Analysis 1, because the slower strain expands faster than its single-strain expansion speed its density bulge, while preserving the spatial structure, is depleted exponentially over time. This is expressed mathematically as \( \rho_{\text{slow}} (y, t) = \rho_{\text{slow}}^* (y - dt^*) e^{-dt^*} \) where \( \rho_{\text{slow}}^* \) also has a bulge for \( y > 0 \) and a trailing exponential for \( y < 0 \). Because the density bulges of the two strains are aligned spatially by the common attractant gradient, we denote the position of the bulge peak by \( y^* \) for both strains. However, in general neither the peak height nor the slope of the trailing exponentials would be the same between the two strains. In b, the density profile of the slower strain, \( \rho_{\text{slow}}^* (y) e^{-dt^*} \), is illustrated at three times \( t \). Dotted purple line indicates initial time \( t = 0 \) where the density bulge of the slower strain has the same peak value \( \rho^* \) as the faster strain (green line): \( \rho_{\text{slow}} (y^*) = \rho^* \). Dashed purple line indicates density profile of the slower strain at time \( t \), where the two density troughs cross: \( \rho_{\text{slow}} (0) e^{-dt^*} = \rho_0 \). The location at which the troughs cross (open red circle) is the crossover distance \( d = u_{\text{fast}} t^* \) (obtained from \( y(t^*) = 0 \)). For \( t > t^* \), the density bulge of the slower strain sinks steadily below that of the faster strain, and its density profile (solid purple line) crosses that of the faster strain (green line) at \( y < 0 \) behind the front and falls steadily backward in the moving frame (filled red circle). Supplementary Analysis 1 shows that for a steady sinking rate \( y = (u_{\text{fast}} - u_{\text{slow}}) \), the corresponding crossover position in the laboratory frame remains at \( d \), for all \( t > t^* \) (fixed position of red circle in b). This feature of the crossover dynamics can be understood intuitively: as chemotaxis occurs only within the bulge region, cells at the crossover point (and to its left) experience no net drift. As the two strains grow at the same rate, their densities at this position remain equal to each other for all \( t > t^* \), implying that the density crossover is frozen in at \( d \). The mathematical analysis provides an expression (Supplementary equation (E14); see Supplementary Information) of how the crossover distance \( d = u_{\text{fast}} t^* \) depends on \( u_{\text{fast}} \) and \( u_{\text{slow}} \) and on static properties of the two density profiles, \( \rho_{\text{fast}}^* \) and \( \rho_{\text{slow}}^* \). From this, we can obtain an expression for the stability distance, defined in the limit the two speeds approach each other: \( d^* = u_{\text{fast}} t^* \). Supplementary Analysis 2 shows that \( t_{\text{fast}} (u_{\text{fast}} - u_{\text{slow}}) \) approaches a finite limit that is proportional to \( 1/\lambda \), the doubling time. The proportionality of \( d^* \) and \( \lambda \) is verified in Extended Data Fig. 6d, and the growth rate dependence of the slope in Extended Data Fig. 6e.
Extended Data Fig. 8 | Fitness landscapes for the competitive expansion process at different positions. The position-dependent fitness landscapes for the competition between two strains (strains 1 and 2, for example, an ancestor and a mutant), with expansion speeds $u_1$ and $u_2$, respectively, computed according to the competitive expansion model in Extended Data Fig. 6a. The fitness of the mutant (strain 2) relative to the ancestor (strain 1) is defined as the ratio of its density with respect to the ancestor:

$$W_X(\rho) = \lim_{\rho X(t) \to \infty} \rho X(t) \rho X'_t$$.

In practice, we take $t = 24$ h, which is well after the dynamics have halted. We further normalize this fitness by the ratio of the initial inoculant, $W_0 = \rho X(\rho X_t(0))$, to obtain the 'relative fitness gain' $W/W_0$. In $a$, green indicates a gain in the fitness of the mutant ($W/W_0 > 1$) and blue indicates a loss ($W/W_0 < 1$). Top, middle, and bottom rows refer to results at distances $X = 20, 15$ and $10$ mm, respectively.

$\text{a, Landscape} W/W_0$ for equal abundance (50:50) of the two strains at equal inoculation ($W_0 = 1$). Cyan and dashed represent $W/W_0 = 1$ and $u_1 = u_2$, respectively. For $W_0 = 1$, the cyan line is the crossover distance: $X = d(u_1, u_\text{c})$ (corresponding to $W = 1$) (Fig. 4). The middle row ($X = 15$ mm) is an expanded view of Fig. 4c, except that the phase diagram in Fig. 4c shows only qualitative information on which strain dominates, whereas the fitness landscape here also provides quantitative information on the fitness gain. Top and bottom, corresponding fitness landscapes for $X = 20$ and $10$ mm, respectively.

$\text{b, Fitness landscape calculations for a mutant:ancestor inoculant ratio of 5:95 } (W_0 = \rho X(0); p(X, 0) = 5:95)$. The solid brown line still indicates the phase boundary $W/W_0 = 1$, but it is no longer directly related to the crossover distance (see below). The topological features of the landscapes in $a$ and $b$ are similar but the details differ, reflecting the frequency-dependent nature of the competition process.

$\text{c, Landscape profiles for three special ancestor expansion speeds, one at each position} X \text{ indicated by dotted grey lines marked as } ① , ② \text{ and } ③ \text{ in } b \text{, each obtained as the intersection of the phase boundary and the diagonal (solid brown and dashed pink lines, respectively).}$

For $X = 15$ mm (middle), ancestral speed $u_1^{\ast} = 5.15$ mm h$^{-1}$ marked by grey line $②$, the fitness landscape has a maximum at $u_1^{\ast}$ (white circle), indicating that ancestral expansion speed $u_1^{\ast}$ is stable to invasion by mutants with smaller or larger expansion speeds. Similarly, for $X = 20$ mm (top), the ancestral speed $u_1^{\ast} = 6.77$ mm h$^{-1}$ (①) is the stable expansion speed, whereas for $X = 10$ mm (bottom), the ancestral speed $u_1^{\ast} = 3.63$ mm h$^{-1}$ (③) is the stable expansion speed.

$\text{d, The phase boundaries shown as solid cyan and brown lines in } a \text{ and } b \text{ intersect the diagonal at the same speed } u^{\ast} \text{ for each } X$. Thus, the stable expansion speeds can be obtained from the crossover distance (cyan line) based on 50:50 inoculant (equation (3)), despite the frequency dependence of the competition process. In other words, even though the overall fitness $W$ depends on the initial frequency $W_0$ in a complex way, in the vicinity of the diagonal ($u_1 = u_2$), $W$ is independent of $W_0$, so that the illustration in Fig. 3a of the stability of a strain of speed $u$ at distance $d(u, u)$ is verified explicitly in $c$. To vary expansion speeds in each panel, the chemotactic coefficients $c_i$ were varied within the range $180-1,500 \mu m^2 s^{-1}$. Diffusion coefficients were changed accordingly ($D = c_i/6.25$; see Supplementary Model and Supplementary Table 3).
Extended Data Fig. 9 | Experimental evolution of expanding bacterial populations in various agar concentrations. a, Three lineages of evolution trajectories in semi-solid TB plates with different agar concentrations. b, c, Illustration of how stability distance is determined for each ancestral expansion speed based on the data shown in a. Using linear regression of the mean evolution trajectories obtained (here for 0.25% agar, ancestral ES = 6 mm h⁻¹, n = 3 biological replicates as in a), we obtained a slope for each series, A–E, as in b. Then we plotted the slope obtained in b against the corresponding selection position to obtain another line (c). The stability distance is estimated as the x intercept of the line. d, Batch culture growth curve of ancestor cells grown in TB (purple) or CAA (brown) medium; the growth rate in CAA is almost 50% slower. Two replicates showed similar results. e, The evolution experiment in CAA medium for five agar densities (same as those shown for TB in a). Three independent lineages were run in parallel as for TB.
Extended Data Fig. 10 | Stable expansion speeds in spatial habitats of different sizes. **a**, Illustration of a recurring ecological scenario. A mixed community of species with different motility characteristics expands into an enclosed habitat of a certain size. All cells within the habitat have an equal chance of being passaged on to occupy a new habitat of the same size. The competitive expansion model (Extended Data Fig. 6a) was extended to model the co-expansion of five distinct species, chemotaxing on the same attractant gradient (Supplementary Analysis 5). Each species has a distinct expansion speed when expanding alone (between 2 mm h⁻¹ and 8 mm h⁻¹), due solely to different motility characteristics that are modelled by different chemotactic coefficients \( c_i \) in the range \( 34 \ldots 274 \, \mu m^2 s^{-1} \). A closed (zero-flux) boundary condition is applied so that when cells reach the edge of the habitat, they cannot propagate forward but they can propagate backward by diffusion (backward propagation via chemotaxis along reversed chemoattractant gradients is possible in principle but very limited, because the chemoattractant is mostly consumed when the population expands forward throughout the habitat). **b**, Average expansion speed of the population changes over different cycles of the simulated expansion–selection process. The average expansion speed increases as the selection proceeds for the largest habitat size (25 mm, pink) and decreases for the smallest habitat size (10 mm, green). The relative abundance of species with different single-strain expansion speeds is changed after each round of selection. The result after the fifth round is shown on the right for the three habitat sizes shown. The faster species is enriched in the largest habitat (top) and the slower species is enriched in the smallest habitat (bottom). The slower species take more cycles to emerge as the winner since they occupy the habitat interior and receive lower weights owing to two-dimensional geometry. **c, d**, To understand why the faster species were selected against in the smaller habitat, we plotted the radial density function \( \rho(r,t) \) for each species \( i \) at \( t = 10 \) h and 24 h after expansion during the first selection cycle (before any selection took place). The density profiles show that after the faster species reached the edge of the habitat, they moved backward towards the interior. However, the speed of this backward movement was limited compared to the outward movement. This is attributed to the fact that the chemoattractant behind the front is depleted, so the backward movement cannot rely on chemotaxis but is a result of cell diffusion (via Komolgorov–Fisher dynamics). In **d**, the cyan and blue lines (expansion speeds 7 and 7.8 mm h⁻¹, respectively) almost overlap and are hard to distinguish visually.
## 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](https://www.nature.com/authors) and the [Editorial Policy Checklist](https://www.nature.com/policies).

### Statistics

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

| Item                                                                 | Confirmed |
|---------------------------------------------------------------------|-----------|
| The exact sample size \( n \) for each experimental group/condition, given as a discrete number and unit of measurement | X         |
| A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly | X         |
| The statistical test(s) used AND whether they are one- or two-sided | X         |
| Only common tests should be described solely by name; describe more complex techniques in the Methods section. |           |
| A description of all covariates tested                              | X         |
| 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) | X         |
| For null hypothesis testing, the test statistic (e.g. \( F \), \( t \), \( r \)) with confidence intervals, effect sizes, degrees of freedom and \( P \) value noted | X         |
| Give \( P \) values as exact values whenever suitable.            | X         |
| For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings | X         |
| For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes | X         |
| Estimates of effect sizes (e.g. Cohen’s \( d \), Pearson’s \( r \)), indicating how they were calculated | X         |

Our web collection on [statistics for biologists](https://www.nature.com/articles/nmeth.4413) contains articles on many of the points above.

### Software and code

Policy information about availability of computer code

**Data collection**

| Imaging data was collected using the software NIS-Elements AR 4.50 (Nikon Ti-E Microsystems). The expansion speed data was collected using a Canon EOS 600D digital camera. The cell counting data was collected by software CytExpert2.1 (Beckman, Cyto-FLEX). The quantitative real-time RT-PCR data was collected by Bio-rad CFX and software Bio-rad CFX manager 3.1. Single cell tracking data was collected by Nikon Ti-E and MicroManager1.4. All the sequencing results were collected using a BGI Illumina system (HiSeq X10). |

**Data analysis**

| NIS-Elements AR 4.50 was used to analyze the fluorescence intensity profile data. Image J and MATLAB(2019a) were used to analyze the expansion speed data and motility data, CytExpert2.1 was used to analyze the cell counting data, Bio-rad CFX manager3.1 was used to analyze the quantitative real-time RT-PCR data. Sequencing results were analyzed with the bresq pipeline (bresq v0.31.1). GraphPad Prism6 was also used to analyze the growth rate and expansion speed. Custom-made codes used for performing simulations is available at https://github.com/jonascremer/chemotaxis_simulation. |

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

All relevant data are included as source data and are available from the corresponding author on request. Sequencing data were deposited to the NCBI Sequence Read Archive(SRA) with the accession PRJNA559221. Data are available at https://www.ncbi.nlm.nih.gov/sra/PRJNA559221. Custom-made code used for performing simulations is available via GitHub at https://github.com/jonascremer/chemotaxis_simulation.
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 | No statistical methods were used to predetermine sample size. Sample size was determined according to the minimal number of independent biological replicates that significantly identified an effect. |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data exclusions | No data was excluded from the analysis.                                                                                                                                                    |
| Replication | Repeated measurements of the evolving quantities (expansion speeds, growth rates, relative fitness, profiles and mean run speed/tumble bias) showed deviations of less than 10% confirming replication of the reported experiments. |
| Randomization | Not applicable.                                                                                                                                             |
| Blinding | Data collection followed the same predetermined protocols throughout the whole study, so no blinding was performed. The custom made codes for the analysis of density profiles (to derive fitness) and the analysis of swimming behavior did not include any adjustable parameters and thus did not require blinding. |

Behavioural & social sciences study design

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

| Study description | Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study). |
| Research sample | State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source. |
| Sampling strategy | Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed. |
| Data collection | Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection. |
| Timing | Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort. |
| Data exclusions | If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established. |
| Non-participation | State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation. |
| Randomization | If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled. |

Ecological, evolutionary & environmental sciences study design

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

| Study description | Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates. |
| Research sample | Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and |
any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy
Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection
Describe the data collection procedure, including who recorded the data and how.

Timing and spatial scale
Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken.

Data exclusions
If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Reproducibility
Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

Randomization
Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Blinding
Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work?  
☐ Yes  ☐ No

Field work, collection and transport

Field conditions
Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

Location
State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

Access and import/export
Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

Disturbance
Describe any disturbance caused by the study and how it was minimized.

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 |

Antibodies

Antibodies used
Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.

Validation
Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer’s website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.

Eukaryotic cell lines

Policy information about cell lines
State the source of each cell line used.
## Palaeontology

| Specimen provenance | Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). |
|---------------------|------------------------------------------------------------------------------------------------------------------------|
| Specimen deposition | Indicate where the specimens have been deposited to permit free access by other researchers. |
| Dating methods      | If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement, where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided. |

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

## Animals and other organisms

### Policy information about studies involving animals: ARRIVE guidelines

| Laboratory animals | For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals. |
|--------------------|------------------------------------------------------------------------------------------------------------------------|
| Wild animals       | Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals. |
| Field-collected samples | For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field. |
| Ethics oversight   | Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Human research participants

### Policy information about studies involving human research participants

| Population characteristics | Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above." |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------|
| Recruitment                 | Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results. |
| Ethics oversight            | Identify the organization(s) that approved the study protocol. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

### Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

| Clinical trial registration | Provide the trial registration number from ClinicalTrials.gov or an equivalent agency. |
|----------------------------|------------------------------------------------------------------------------------------------------------------------|
| Study protocol             | Note where the full trial protocol can be accessed OR if not available, explain why. |
| Data collection            | Describe the settings and locales of data collection, noting the time periods of recruitment and data collection. |
| Outcomes                   | Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures. |
ChIP-seq

Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as GEO.
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session (e.g. UCSC)

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

- 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

Samples were firstly fixed with pre-cooled cell counting buffer (0.9 % NaCl with 0.12 % formaldehyde). Then the fixed samples were diluted as necessary with staining buffer (cell counting buffer with 0.1 μg/ml DAPI) prior to the flow cytometer analysis. Details are provided in the Methods.

Instrument

Beckman, Cyto-FLEX

Software

CytExpert2.1

Cell population abundance

At least 50,000 cells per sample were analyzed.

Gating strategy

DAPI staining was applied to distinguish bacterial cells from other particles. DAPI-stained particles were deemed as the bacterial cells, and the purity was above 95%. Cells were separated into two groups GFP positive/GFP negative through the FITC channel. Obtained results agreed with fluorescence intensity measurements using microscopy and scanning the entire population.

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

**Design type**
- Indicate the task or resting state; event-related or block design.

**Design specifications**
- Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

**Behavioral performance measures**
- State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

### Acquisition

**Imaging type(s)**
- Specify: functional, structural, diffusion, perfusion.

**Field strength**
- Specify in Tesla

**Sequence & imaging parameters**
- Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

**Area of acquisition**
- State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

**Diffusion MRI**
- Used
- Not used

### Preprocessing

**Preprocessing software**
- Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

**Normalization**
- If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

**Normalization template**
- Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

**Noise and artifact removal**
- Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

**Volume censoring**
- Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

### Statistical modeling & inference

**Model type and settings**
- Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects, drift or auto-correlation).

**Effect(s) tested**
- Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

**Specify type of analysis:**
- Whole brain
- ROI-based
- Both

**Statistic type for inference**
- Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

(See Eklund et al. 2016)

**Correction**
- Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

### Models & analysis

**n/a**
- Involved in the study

- Functional and/or effective connectivity
- Graph analysis
- Multivariate modeling or predictive analysis

**Functional and/or effective connectivity**
- Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

**Graph analysis**
- Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).
Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.