Cell Reports, Volume *18*

Supplemental Information

Evolutionary Remodeling of Bacterial Motility Checkpoint Control

Bin Ni, Bhaswar Ghosh, Ferencz S. Paldy, Remy Colin, Thomas Heimerl, and Victor Sourjik

Supplemental Information

Evolutionary Remodeling of Bacterial Motility Checkpoint Control

Bin Ni, Bhaswar Ghosh, Ferencz S. Paldy, Remy Colin, Thomas Heimerl, Victor Sourjik

SUPPLEMENTAL FIGURES

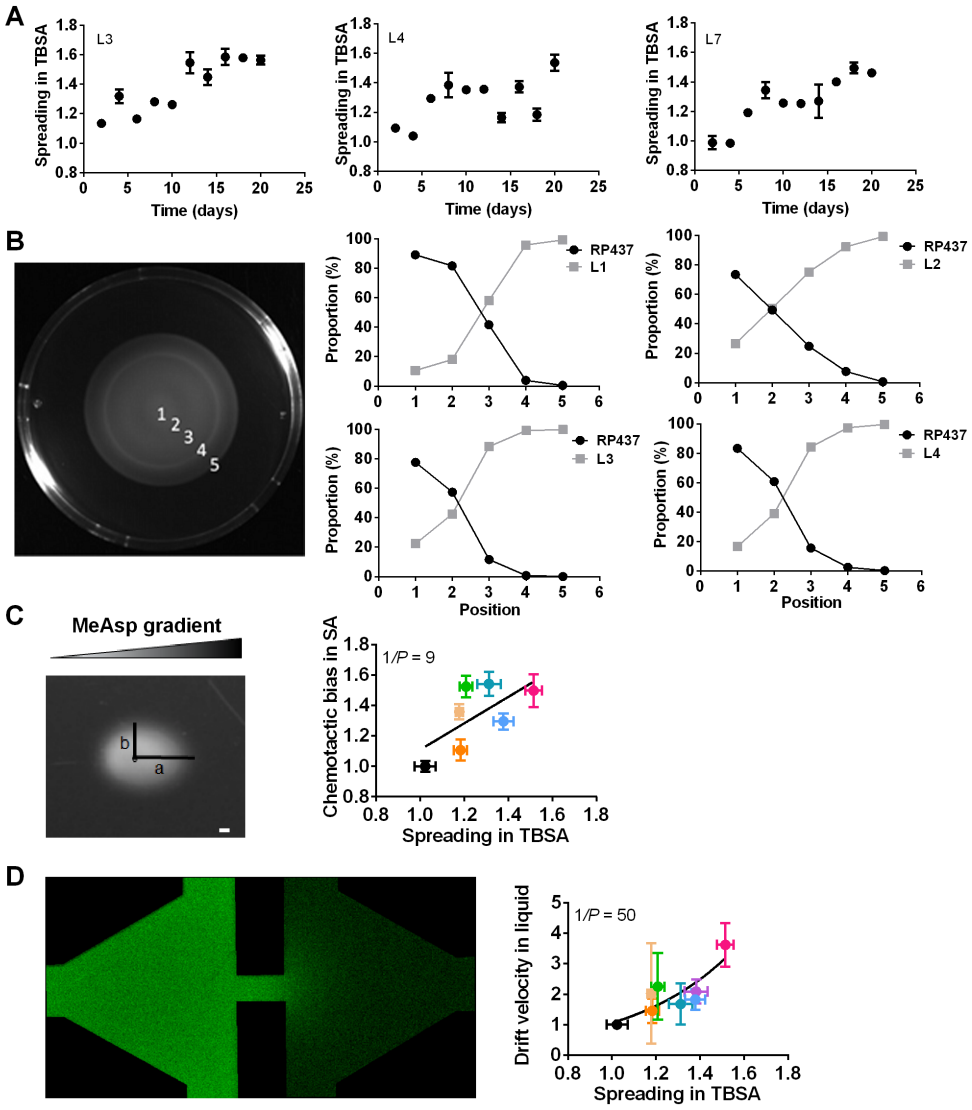

Figure S1. Experimental Evolution of *E. coli* Spreading and Its Characterization. Related to Figure 1. (A) Examples of time course of evolution of spreading in TBSA for lines L3, L4 and L7. **(B)** Competition assay in spreading between evolved strains and reference parental strain RP437. Evolved strains (labeled with CFP; pVS129) were co-inoculated 1:1 with RP437 (labeled with YFP; pVS132). Samples were collected at indicated positions along the colony spreading in TBSA (left panel) and fractions of both strains at each position were quantified using flow cytometry (right panels). Data in (A) and (B) are represented as mean ± SEM of three replicates. **(C)** Correlation between spreading in TBSA and biased movement (chemotactic bias) in minimal medium SA with pre-established gradient (0 - 100 μM) of MeAsp. Chemotactic bias was quantified as a ratio of spreading up the gradient to the spreading normal to the gradient (a/b). Scale bar is 1 mm. **(D)** Correlation between spreading in TBSA and chemotactic drift velocity in MeAsp gradient established in a liquid-filled PDMS chamber. Gradient formation in the microfluidic device was monitored using fluorescein. The channel dimensions are 2 x 1 mm. Correlation values are 0.65 (C) and 0.78 (D). Line in (D) indicates an exponential fit.

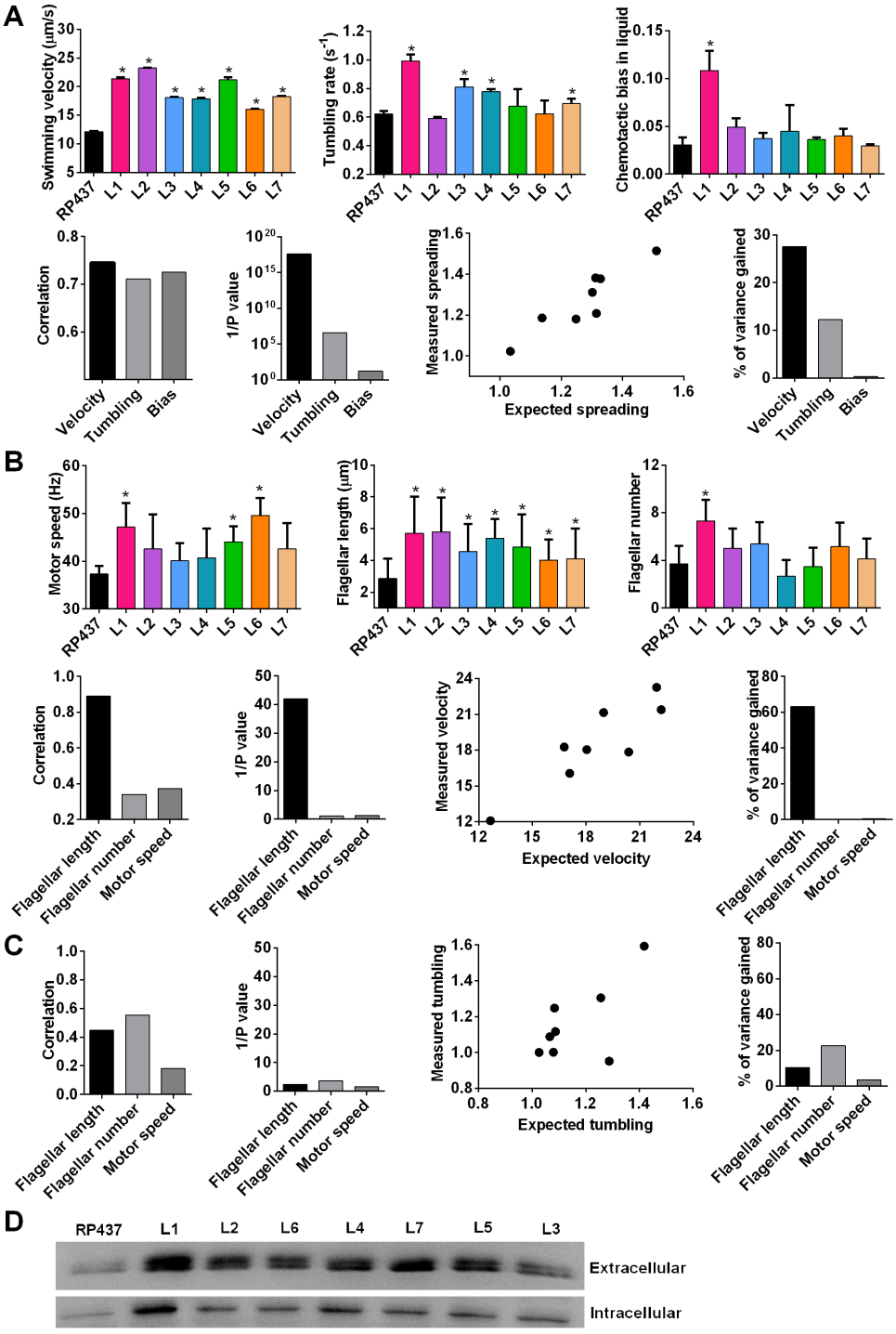

Figure S2. Regression Analysis of Correlation between Spreading, Motility and Flagellation in Evolved Strains. Related to Figure 2. (A) Correlation between spreading and velocity, tumbling rate and bias for motility analysis performed in liquid in absence of chemical gradients. Chemotactic bias in liquid was defined as chemotactic drift (Fig. S1D) normalized by cell swimming velocity. Asterisks indicate significant differences from PR437 according to two-tailed *t*-test ($P < 0.05$). Regression analysis was performed using linear regression model as described in the Supplemental Experimental Procedures. **(B)** Correlation between velocity and flagellar length, flagellar number and motor speed.

Rotation rate of individual flagellar motors in evolved cells was determined by tracking 1.1 μm beads attached to flagella. Electron microscopy was used to quantify flagellar length and number in evolved strains. **(C)** Correlation between tumbling rate and flagellar length, flagellar number and motor speed. **(D)** Western Blot analysis of flagellin levels inside and outside the cell for RP437 and evolved strains.

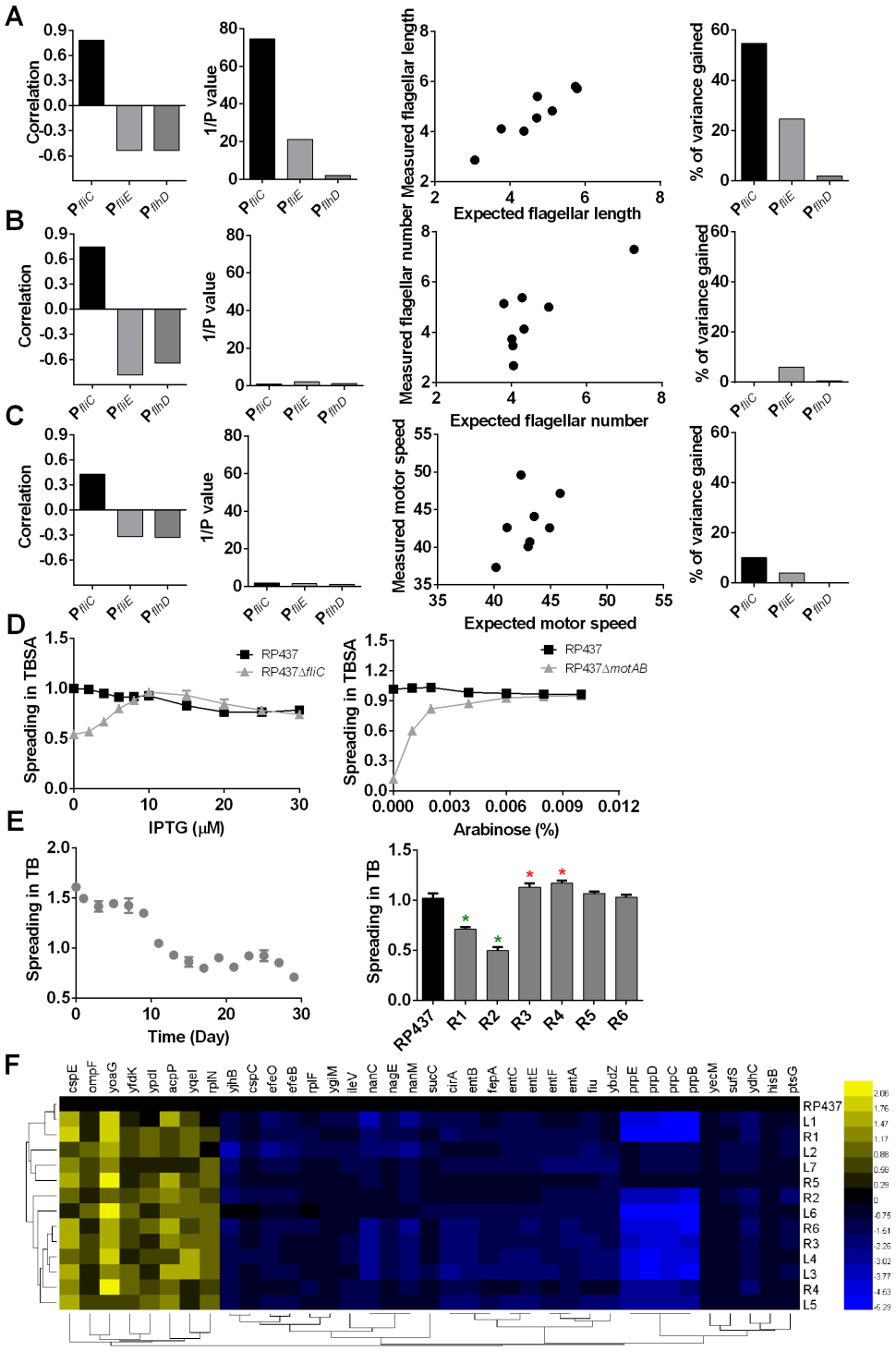

Figure S3. Changes in Gene Expression during Evolution. Related to Figure 2. (A-C) Regression analysis of correlation between flagellar length (A), flagellar number (B) or motor speed (C) and P_{fliC}, P_{fliE} or P_{flhD} promoter activities. Regression analysis was performed using linear regression model as described in the Supplemental Experimental Procedures. **(D)** Titration of FliC (left) or MotA and MotB (right) either in RP437 or in corresponding knockout strains. **(E)** Example of time evolution of spreading efficiency for L1 strain grown in liquid TB (left) and spreading efficiency after 20-30 rounds of selection for individual lines of evolution (right). Asterisks indicate significance for decrease (green)

or increase (red) in spreading compared to RP437 according to *t*-test ($P < 0.05$). Data in (D and E) are represented as mean ± SEM of three replicates. **(F)** Genes that changed their transcription independent of selection for motility according to the RNAseq analysis.

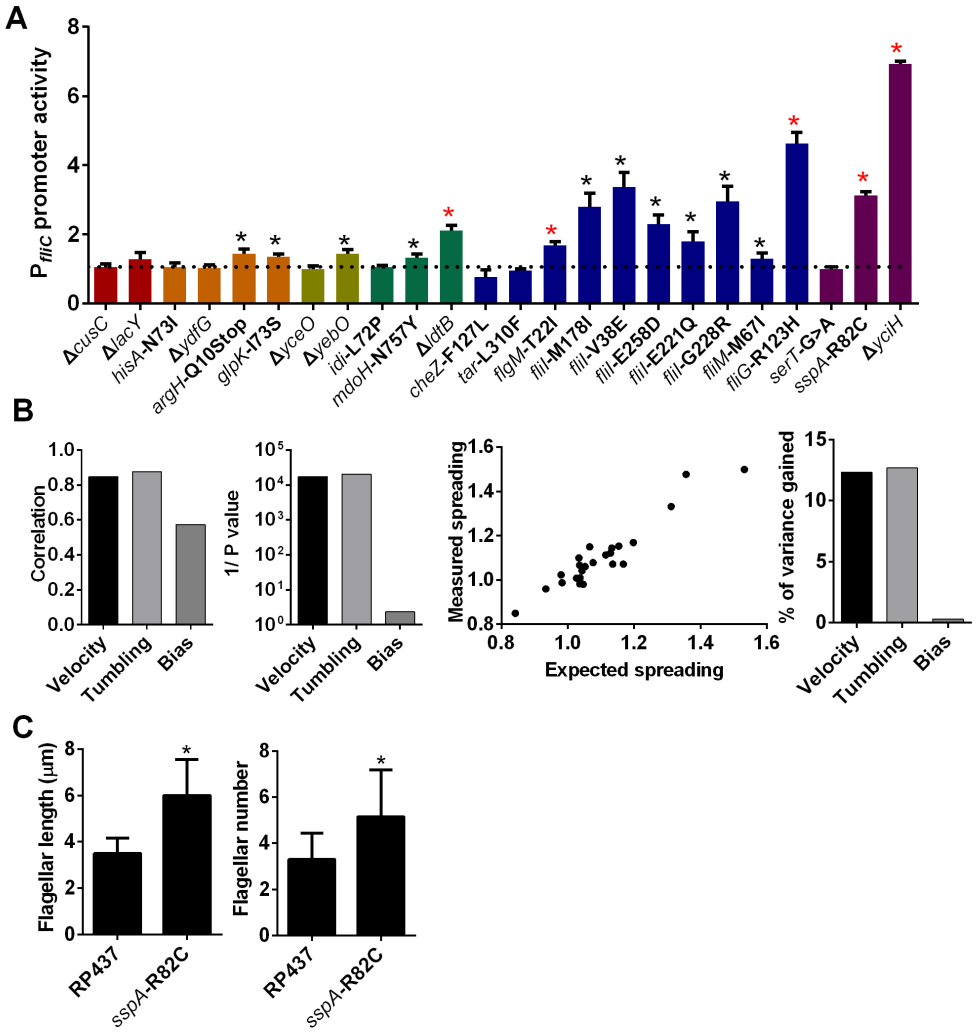

Figure S4. Analysis of Effects of Individual Mutations. Related to Figure 3. (A) P_{fliC} promoter activity for RP437 strain carrying individual adaptive mutations, normalized to RP437. Significant increase (black asterisks, $P < 0.05$; red asterisks, $P < 0.01$) in expression compared to the parental strain RP437 is indicated. **(B)** Regression analysis of correlation between spreading in TBSA and velocity, tumbling and chemotactic bias. Regression analysis was performed using linear regression model as described in the Supplemental Experimental Procedures. **(C)** Effects of R82C mutation in *sspA* on flagellar length and number. Asterisks indicate significant increase in number and length of flagella compared to RP437 according to *t*-test ($P < 0.05$). Data in (A and C) are represented as mean ± SEM of three replicates.

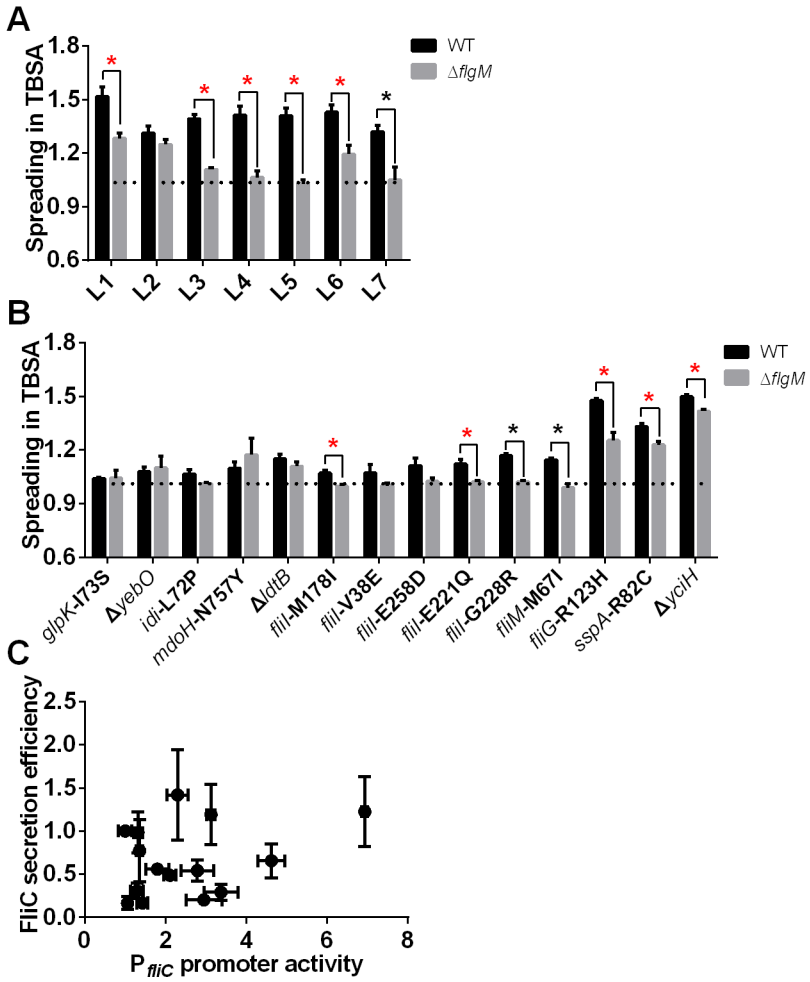

Figure S5. FlgM Secretion Controls on Spreading in TBSA. Related to Figure 4. (A,B) Spreading of evolved strains (A) or of RP437 carrying individual mutations (B), either in presence of FlgM (black bars) or in *flgM* knockout strain (grey bars). In each case, spreading was normalized to the respective reference strain, RP437 or RP437Δ*flgM*. Asterisks indicate significant difference between relative spreading in $FlgM^+$ and Δ*flgM* backgrounds (black: $P < 0.05$; red: $P < 0.01$) according to two-tailed *t*-test. (C) Correlation between FliC secretion efficiency and P_{fliC} promoter activity in strains containing individual mutation.

SUPPLEMENTAL TABLES

Table S1. Mutations Found in Evolved Strains by Genome Resequencing. Related to Figure 3.

Function	Genes	Annotation	L1	L2	L3	L4	L5	L6	L7
Motility	*fliI*	ATPase of flagellar export apparatus	E258D	M178I		G228R	V38E		E221Q
	fliM	flagellar motor switch protein	M67I						
	fliG	flagellar motor switch protein				R123H			
	tar	chemoreceptor		L310F					
	cheZ	phosphatase for CheY-P			F127L				
	flgM	anti-sigma factor for FliA			T22I				
Regulatory factor	*serT*	serine tRNA	C>T[a]						
	sspA	stringent starvation protein A	R82C						
	yciH	translation initiation factor		Q36Stop					
Cell envelope biogenesis	*ldtB*	L, D-transpeptidase						IS1	
	mdoH	membrane glycosyltransferase		N757Y					
	idi	isopentenyl diphosphate isomerase							L72P
Metabolism	*hisA*	L-histidine biosynthesis	N73I				N73I	N73I	
	argH	L-arginine biosynthesis		Q10Stop					
	glpK	glycerol kinase		I73S					
	ydfG	3-hydroxy acid dehydrogenase		IS1[b]					
Transporter	*cusC*	copper / silver efflux system		IS1					
	lacY	lactose permease	Del						
Unknown	*yebO*	unknown			IS1				
	yceO	possible regulator of biofilm formation / acid stress		IS1	IS1	IS1			IS1

[a]Mutation from guanine to adenine at position 31 in *serT* RNA.

[b]Insertion of transposable IS1 sequence.

Table S2. Mutations in Motility Genes Found in Strains Evolved in Minimal Medium. Related to Figure 7.

Genes	Annotation	Asp_1[a]	Asp_2	Ser_1	Ser_2	Glc_1	Glc_2	MA_1	MA_2
fliH	regulatory component of FliI				Q217H				H95L
fliI	ATPase of flagellar export apparatus	R231L	V301A	T377M		V70L	S106C		
fliM	flagellar motor switch protein							T192N	

[a]All lines were evolved in in Minimal A medium SA with 0.2% glucose as carbon source. Lines Asp_1, Asp_2, Ser_1 and Ser_2 were additionally supplemented with 1 mM aspartate or serine, respectively, as chemoattractant. For lines Glc_1 and Glc_2 glucose was both carbon source and chemoattractant. Lines MA_1 and MA_2 were evolved in a pre-established gradient of MeAsp (see Fig. S1C).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Strains and Plasmid Construction

pKOV was used to integrate single SNP mutation into the genome of wild type strain (Link et al., 1997). Gene deletions were constructed by P1 transduction from KEIO collection (Baba et al., 2006). Kanamycin resistance cassette was removed using pCP20 (Cherepanov and Wackernagel, 1995). GFP reporters for *flhD* and *fliC* promoters were constructed based on pUA66 (Zaslaver et al., 2006), yielding pAM104 and pAM109, respectively. Promoter reporters for other genes were from a pUA66-based promoter library (Zaslaver et al., 2006). pTrc99a vector inducible by isopropyl β-D-1-thiogalactopyranoside (IPTG) was used to express FliA (pBN2), FlgM (pBN3), FlgM fused C-terminally to HA tag (pBN4) or FliC (pBN5). pBAD18-Kan vector inducible by arabinose was used to express MotAB (pBN6). pKAF131 carrying sticky *fliC* allele under native *fliC* promoter (Yuan et al., 2010) was a gift of Karen A. Fahrner and Howard C. Berg. For bead assay, *fliC* knockout strains carrying pKAF131 were used.

Immunoblot Analysis of FlgM or FliC Secretion

Cell cultures were adjusted to the same density (OD_{600} = 0.5-0.6) and centrifuged at 4,000 rpm for 8 min. Proteins in cell free supernatant were precipitated by 25 % trichloroacetic acid (TCA) and incubated on ice for 5 min. Precipitated proteins were collected by centrifugation (10,000 rpm) at 4 °C and washed twice with ice-cold acetone. After drying, 1X sample buffer containing 8 M urea was used to dissolve pellet for SDS-PAGE. Samples were separated on 15 % PAGE and transferred to PVDF membrane using wet blotting. Primary anti-HA antibody (Sigma) at a 1:10,000 dilution and secondary anti-mouse IgG antibody labeled with peroxidase (Sigma) at a 1:5,000 dilution were used for detection.

Regression Analysis

In order to asses relative contributions of individual independent variables on the dependent, we constructed a multilinear regression model, where the dependent variable (y) is assumed to be a linear additive function of the independent variables (x_1, x_2, x_3)

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3$$

Linear least square method was used to calculate the values of coefficients. The value of R^2 gives an estimate of the proportion of variance of the dependent variable predictable from the full mutilinear regression model. Statistical significance of estimating each parameter of the model is measured by calculating the *p*-value of the *t*-statistics. The *p*-value of the parameter is the measure of the corresponding independent variable's relative contribution in predicting the dependent variable. The significance of the full model is calculated by estimating the *p*-value of the *F*-statistics of the full model. Since in the full model it is assumed that the independent variables are uncorrelated, we further constructed linear models where one of the independent parameters (x_1) is removed from the model so that the dependent variable is

$$y = a_0 + a_2x_2 + a_3x_3$$

and calculated R^2. The gain in R^2 by adding the variable is a measure of effect of the variable x_1 in explaining the variance of the dependent variable:

$$\Delta R^2(x_1) = R^2(fullmodel) - R^2(modelafterremovingx_1)$$

The values of ΔR^2 corroborate the significances calculated from the *p*-values of the full model. The calculations were performed using the *lm* function of the software *R*.

Spearman Rank Correlation(ρ)

The association between each pair of variables was calculated using Spearman rank correlation (ρ). The values of each variables were ranked and the correlation between the ranks was calculated. Since rank correlation does not assume any functional relationship between the variables, it is appropriate when the relationship is unknown. The corresponding significance of the correlation value was estimated by Spearman rho test which computes a probability of the correlation value being zero. The calculations were performed by the cor.test function of the software *R*.

Evolutionary Algorithm

Evolutionary simulations for topologies shown in Figure 7B were performed in MATLAB using standard genetic algorithm framework (Mitchell, 1998). Each simulation starts with a homogeneous population of 500 individual networks with the same topology and identical parameter values. In each generation, one of the model parameters is mutated randomly from a uniform distribution from 0 to 1 in each of the individual networks selected by the mutation rate 0.8 (i.e., in 80% of the networks). The I/O behavior of the network is modeled as a linear relationship. Thus, for the linear network with one input (x), one intermediate (y) and one output (z),

$$y = y_T w_1 x$$
$$z = z_T w_2 y$$

Here w_1 and w_2 denote the strength of the links between the input and the intermediate and between the intermediate and the output, respectively; y_T and z_T represent maximal attainable values of y and z, respectively, for one incoming link.

For the linear network with three inputs and three intermediates but one output,

$$y_1 = y_{1T} w_1 x_1; y_2 = y_{2T} w_2 x_2; y_3 = y_{3T} w_3 x_3;$$
$$z = z_T (w_4 y_1 + w_5 y_2 + w_6 y_3)$$

For the bow-tie network,

$$y = y_T (w_1 x_1 + w_2 x_2 + w_3 x_3)$$
$$z = z_T w_4 y$$

The fitness is defined to be proportional to the exponential of the output,

$$F = e^{(0.1+z)}$$

and it is calculated for each individual network in the population. Notably, in our model the maximal attainable fitness of the network is determined by the number of inputs that can acquire mutations, so that it is lower for a linear network with one input but the same for the linear and bow-tie networks with three inputs.

During *in-silico* evolution, the networks for the next generation are selected among the parental networks with replacement following proportional reproduction with the exponential scaling (Lampert and Tlusty, 2009; Burda et al., 2010; Friedlander et al., 2015). The fitness of the whole population is calculated as average fitness over the 500 networks. For each case the simulation was repeated for 100 different initial parameter sets so that evolution starts from a different point in the genotype space in order to avoid dependence on the initial parameter values. Median value of the fitness over these 100 different evolutionary runs was taken as a measure of the time trace of evolution.

SUPPLEMENTAL REFERENCES

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006.0008.

Burda, Z., Krzywicki, A., Martin, O.C., and Zagorski, M. (2010) Distribution of essential interactions in model gene regulatory networks under mutation-selection balance. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *82*, 011908.

Cherepanov, P.P., and Wackernagel, W. (1995). Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene *158*, 9-14.

Friedlander, T., Mayo, A.E., Tlusty, T., and Alon, U. (2015) Evolution of Bow-Tie Architectures in Biology. PLoS Comput. Biol. 11, e1004055.

Lampert, A., and Tlusty, T. (2009) Mutability as an altruistic trait in finite asexual populations. J. Theor. Biol. 261, 414–422.

Link, A.J., Phillips, D., and Church, G.M. (1997). Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. J. Bacteriol. *179*, 6228-6237.

Mitchell, M. (1998). An introduction to genetic algorithms. Cambridge, Massachusetts, MIT Press.

Yuan, J., Fahrner, K.A., Turner, L., and Berg, H.C. (2010). Asymmetry in the clockwise and counterclockwise rotation of the bacterial flagellar motor. Proc. Natl. Acad. Sci. USA *107*, 12846-12849.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G., and Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. Nat. Methods *3*, 623-628.