Design and implementation of a bacterial signaling circuit

Lazova, M.D.

2013

document version
Publisher's PDF, also known as Version of record

Appendix B

Comparative analysis of the chemotaxis system of natural *Escherichia coli* isolates

The chemotaxis system of *Escherichia coli* is a paradigmatic signaling circuit, characterized by multiple experimental and theoretical studies. However, these research efforts have been concentrated on understanding the chemotactic signaling of a limited number of laboratory strains. Here we study the chemotactic behavior of a subset of natural *E. coli* isolates that represent the genotypic diversity of the species. We demonstrate variability in chemotactic performance and swimming speed, even in closely related strains. We compare the nucleotide sequence of the chemotaxis operon *meche* between strains with different chemotactic performance and demonstrate that the evolution of the chemotaxis operon closely matches that of a set of housekeeping genes. We observe genetic variation resulting from mutations and possibly horizontal gene transfer, and find *E. coli* strains with reduced chemoreceptor species composition. We discuss the possibility that a comparison of the chemotactic traits of natural *E. coli* strains could allow us to uncover quantitative features of chemotaxis that have undergone changes in the recent evolution of this species.

Appendix B

B.1. Introduction

The *Escherichia coli* chemotaxis network, which allows bacteria to orient in gradients of chemical stimuli, is a paradigmatic signaling circuit that represents many general features of signaling systems, such as adaptation, signal amplification and wide dynamic range [12,171,241]. Each of these characteristics has been a subject of numerous quantitative modeling, theoretical and experimental studies [35,120,127,128,137,222,225,226,260], and the behavioral consequences of chemotactic signaling has also been evaluated using a number of standard assays [2,171,276] as well as microfluidics technology [7,8,156]. The chemotaxis network, like other biological signaling systems [198], is a subject of evolutionary optimization [126], as exemplified by the finding that clinical and natural isolates of *E. coli* vary in their chemotactic performance [133,262]. The extensive knowledge about the molecular-level mechanisms of the chemotaxis system of *E. coli* present a rare opportunity to study the design principles involved in evolutionary optimization [127], which requires generation of phenotypic variability while preserving the important functional features of the chemotactic signaling. However, the vast majority of the experimental work and the information about the properties of *E. coli* chemotactic signaling and behavior have been derived from only a small number of laboratory strains [104], which do not represent the genetic and ecological diversity of *E. coli* [150,189].

E. coli is ubiquitous in nature, with an estimated total population of ~10^{20} bacteria [273]. Its primary habitat is the vertebrate gut, where it is the predominant aerobic commensal organism [256]. Some *E. coli* strains also represent intraintestinal or extraintestinal pathogens [130,133]. Secondary habitats of *E. coli* are water, sediments and soil [70,110]. The population genetic structure is defined by the balance between recombination and mutation [256]. Low recombination levels define a clonal structure, whereas high recombination levels define a panmictic structure [236]. In the pre-sequencing era, studies using multilocus enzyme electrophoresis (MLEE) analysis [219] suggested that the population structure of *E. coli* is predominantly clonal [173,220]. MLEE analysis of thousands of natural and clinical isolates allowed the assembly of a set of 72 reference strains, the ECOR collection, that represented the known genetic diversity of *E. coli* [189] (Table B.1). The ECOR

collection was subdivided into four groups: A, B1, B2, and D, plus a minor group E [275], where group D diverged first and the sister groups A and B1 separated later [182-184]. More recent studies suggest that group B2 is the ancestral one [79,139].

ECOR strain	Location	Host	MLEE group	Spreading in TB soft-agar	Swimming in motility buffer	Doubling time (h)
1	Iowa	Human (F)	A	L	F	
2	New York	Human (M)	A	L	F	0.67
3	Massachusetts	Dog	A	L	F	
4	Iowa	Human (F)	A	L	F	
5	Iowa	Human (F)	A	L	S	
6	Iowa	Human (M)	B1	NM	NM	
7	Washington zoo	Orangutan	A	L	F	
8	Iowa	Human (F)	A	L	F	
9	Sweden	Human (F)	A	NM	NM	
10	New York	Human (F)	A	L	F	
11	Sweden	Human (F)	A	NM	NM	
12	Sweden	Human (F)	A	L	F	0.73
13	Sweden	Human (F)	A	L	S	0.77
14	Sweden	Human (F)	A	L	F	
15	Sweden	Human (F)	B1	L	S	
16	Washington zoo	Leopard	A	L	S	
17	Indonesia	Pig	A	NM	NM	
18	Washington zoo	Celebese ape	A	D	S	1.05
19	Washington zoo	Celebese ape	A	L	F	0.73
20	Bali	Steer	A	L	F	
21	Bali	Steer	A	L	F	0.74
22	Bali	Steer	A	L	F	
23	Washington zoo	Elephant	A	S	S	
24	Sweden	Human (F)	A	NM	NM	
25	New York	Dog	A	D	NM	
26	Massachusetts	Human Infant	B1	L	F	
27	Washington zoo	Giraffe	B1	L	S	0.67
28	Iowa	Human (F)	B1	NM	NM	
29	Nevada	Kangaroo rat	B1	L	S	0.72
30	Alberta	Bison	B1	L	S	
31	Washington zoo	Leopard	E	L	S	0.67
32	Washington zoo	Giraffe	B1	D	S	0.63
33	California	Sheep	B1	L	F	0.67
34	Massachusetts	Dog	B1	NM	NM	
35	Iowa	Human (M)	D	NM	NM	
36	Iowa	Human (F)	D	NM	NM	
37	Washington zoo	Marmoset	E	D	S	0.59
38	Iowa	Human (F)	D	NM	NM	

Appendix B

39	Sweden	Human (F)	D	NM	NM	
40	Sweden	Human (F)	D	NM	NM	
41	Tonga	Human (M)	D	NM	NM	
42	Massachusetts	Human (M)	E	L	F	
43	Sweden	Human (F)	E	NM	NM	
44	Washington zoo	Cougar	D	L	S	
45	Indonesia	Pig	B1	L	F	0.76
46	Washington zoo	Celebese ape	D	NM	NM	
47	New Guinea	Sheep	D	L	F	
48	Sweden	Human (F)	D	L	F	0.76
49	Sweden	Human (F)	D	S	F	
50	Sweden	Human (F)	D	D	F	0.67
51	Massachusetts	Human infant	B2	L	F	0.75
52	Washington zoo	Orangutan	B2	L	S	
53	Iowa	Human (F)	B2	L	F	
54	Iowa	Human	B2	L	F	
55	Sweden	Human (F)	B2	L	F	
56	Sweden	Human (F)	B2	L	F	
57	Washington zoo	Gorilla	B2	S	S	
58	Washington zoo	Lion	B1	L	F	
59	Massachusetts	Human (M)	B2	L	F	
60	Sweden	Human (F)	B2	D	S	
61	Sweden	Human (F)	B2	L	F	
62	Sweden	Human (F)	B2	S	F	0.55
63	Sweden	Human (F)	B2	NM	NM	
64	Sweden	Human (F)	B2	NM	NM	
65	Washington zoo	Celebese ape	B2	L	F	
66	Washington zoo	Celebese ape	B2	L	F	
67	Indonesia	Goat	B1	L	F	
68	Washington zoo	Giraffe	B1	S	S	0.66
69	Washington zoo	Celebese ape	B1	NM	NM	
70	Washington zoo	Gorilla	B1	NM	NM	
71	Sweden	Human (F)	B1	D	M	
72	Sweden	Human (F)	B1	L	S	

Table B.1. Standard reference collection of *E. coli* strains (ECOR collection) [189]. Color-coding represents phylogenetic groups A (green), B1 (red), B2 (yellow), D (blue), and E (black) [275]. Strains are categorized according to their spreading as non-motile (NM), strains spreading diffusely (D), strains forming small but distinct rings (S), and strains forming large rings (L) (see text). Strains are categorized according to their swimming speed as non-motile (NM), slow (S; swimming speeds <10 μm/s), and fast (F; swimming speeds ≥10 μm/s) (see text). The doubling time (h) of a selected subset of ECOR strains in TB at 33.5° C is shown at the last column.

The subsequent development of DNA sequencing techniques led to the discovery that *E. coli* populations are actually seldom clonal, with frequent recombination events revealing themselves as clustered third-base substitutions [174]. The genetic diversity among members of the ECOR collection was attributed to recombination rather than to mutation [95]. Phylogenetic analysis of concatenated sequences of ECOR housekeeping genes revealed the existence of four clades, largely in agreement with the MLEE groups A, B1, B2 and D, although hybrid groups such as AxB1 and ABD, a result of recombination, also exist within the global *E. coli* population [275]. Thus *E. coli* is more diverse than previously appreciated, although some data suggest population contractions and bottlenecks reduced the diversity 10-30 million years ago [275]. During population expansions in the last five million years the descendants of the four major lineages A, B1, B2 and D have become predominant and represent the majority of the contemporary strains [275].

Here we study the diversity of *E. coli* with respect to their chemotaxis and motility. To largely cover the existing natural diversity in behavior, we selected a subset of ECOR strains, which represent the variations in chemotactic spreading and swimming motility in soft-agar plates. The selected strains contain representatives of each of the phylogenetic lineages of *E. coli*, and also cover strains from different geographic locations, animal and human hosts, as well as pathogens and commensals. We show variability of the chemotactic performance even in strains that are closely related phylogenetically. We observe mutation- and recombination-based changes in the chemotaxis operon *meche*, and compare the genetic diversity of this chemotaxis operon with that of a subset of housekeeping genes. We discuss how further physiological studies on chemotactic signaling in the selected subset of strains could reveal which features of *E. coli* chemotactic signaling and behavior are conserved, and which features have undergone changes in the recent past.

Appendix B

B.2. Diversity of soft-agar plate phenotypes of natural *E. coli* isolates

To observe the natural variations of chemotactic performance of *E. coli*, we tested the spreading of all 72 strains from the ECOR collection, using the classical soft-agar assay for chemotaxis [2] (Figure B.1). Bacteria were inoculated in tryptone-broth (TB) soft-agar plates, and incubated at 30° C. Growing bacteria consume nutrients in the media and the motile and chemotactic strains swim outwards in concentric bands, which form due to chemotactic responses to the spatial gradients, resulting from metabolic consumption. In TB (the nutrient tryptone is a casein hydrolysate), the first band consumes serine and most of the oxygen, the second band aspartate, and the third band threonine [2,276]. Motile but non-chemotactic strains also spread outwards due to bacterial growth and motility, but the spreading is slower and do not produce distinct bands [276]. Non-motile strains only expand around the inoculation point due to growth of the bacteria.

Figure B.1A and B.1B show the spreading of the 72 strains of the ECOR collection after 4h and 7.5 h of incubation, respectively. ~25% of the ECOR strains do not spread even after prolonged incubation (Table B.1), suggesting that they are immotile, which was confirmed after performing phase-contrast microscopy of these strains in liquid medium (Materials and methods). Non-motile strains were observed in each of the phylogenetic groups, with the lowest percent in group B2 (13%) and the highest percent in group D (58%). The rest of the strains spread in TB soft-agar plates and swim in liquid medium; however, both the spreading patterns and the motility strongly differ between strains. Spreading strains either spread diffusely or form most often two rings at the time of observation (Table B.1). The swimming speed varies between a few microns per second and >30 μm/s (Table B.1 shows a rough approximation of the swimming speed for the strains of the ECOR collection, see Section B.3). Variations in the spreading patterns and swimming speeds are observed in each of the phylogenetic groups, although there are more fast-spreading strains in groups A and B1 than in B2 and D.

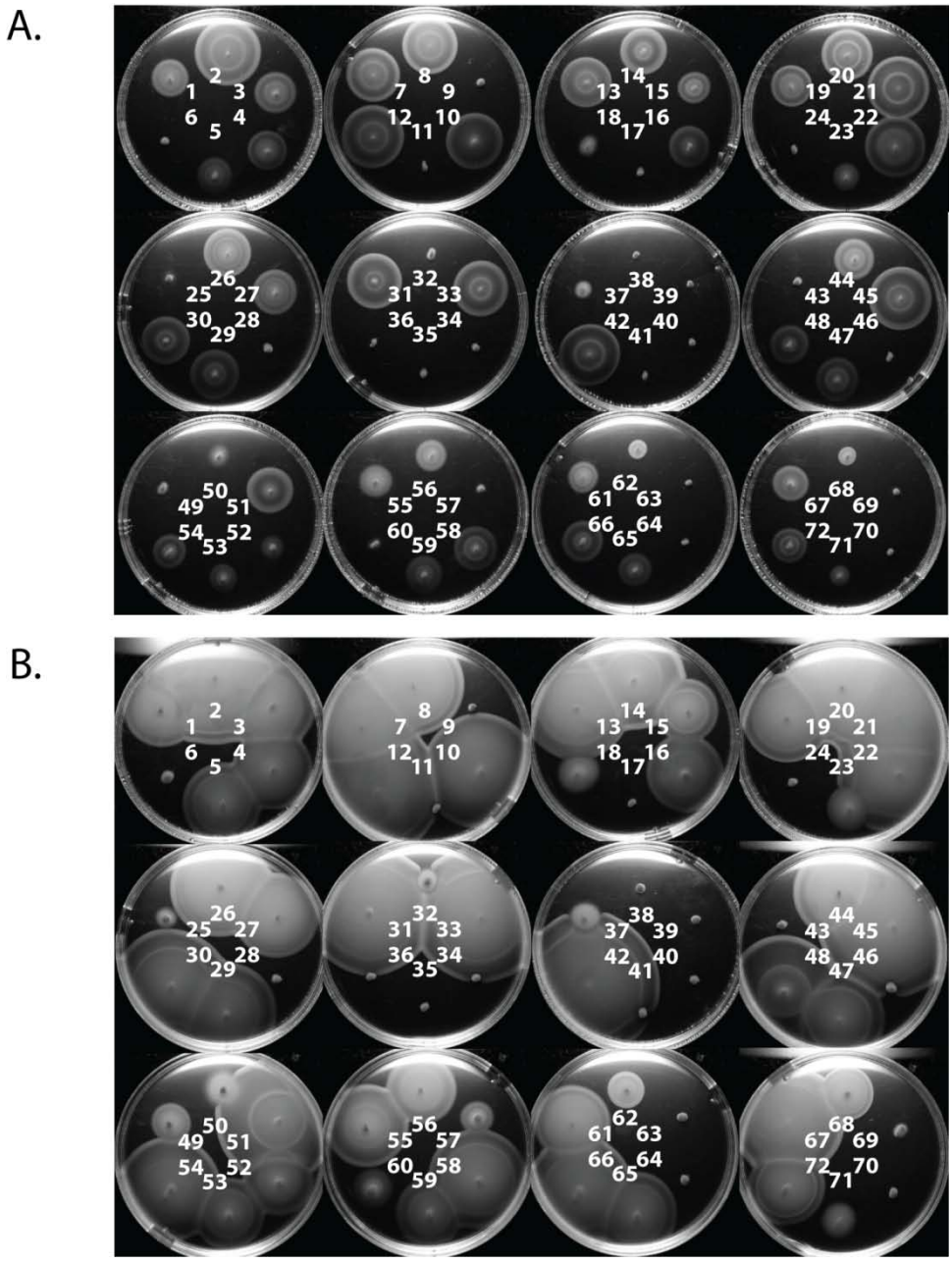

Figure B.1. Soft-agar assays of all 72 strains of the ECOR collection. The numbers correspond to the numbers of the strains in the ECOR collection. The plates were incubated at 30° C and imaged after 4 h **(A)** and 7.5 h **(B)**.

Appendix B

B.3. *E. coli* isolates with distinct chemotaxis characteristics and motility

To analyze the observed diversity of spreading pattern, accounting for the contributions of chemotaxis and other factors, such as differences in growth and random motility [276], we selected for further characterization a quarter of the ECOR strains that show different spreading and belong to different phylogenetic groups and measured their growth in TB and swimming speed in liquid medium. Three groups of strains were selected based on their spreading pattern: strains forming large rings, small but distinct rings, and strains that spread diffusely in TB soft agar. The 18 selected strains have comparable growth rates in TB at 33.5° C with an average doubling time of 0.7±0.1 h (Table B.1). Thus, the observed differences in spreading in TB soft agar are not due to differences in growth, but other factors: mostly chemotactic performance and random motility. Swimming speeds, however, differed between the strains (Table B.1): we labeled the strains to be "fast" if they swim at ≥10 μm/s, and "slow" if they swim at <10 μm/s. Table B.2 summarizes the characteristics of the selected subset of ECOR strains.

ECOR strains	Large rings	Small rings	Diffuse spreading
Fast swimming	**2**, **12**, **19**, **21** **33**, **45** **48** **51**	**62**	**50**
Slow swimming	**13** **27**, **29** **31**	**68**	**18** **32** **37**

Table B.2. Selected subset of ECOR strains, grouped by their spreading patterns and swimming speed. The doubling times in TB at 33.5° C of the selected strains are comparable. The MLEE groups are color-coded: A (green), B1 (red), B2 (yellow), D (blue), and E (black).

B.4. Sequence analysis of the *meche* operon

The canonical chemotactic signaling system of *E. coli* consists of six soluble chemotaxis proteins: the kinase CheA, the scaffolding protein CheW, the adaptation enzymes CheR (methyltransferase) and CheB (methylesterase / deamidase), the response regulator CheY and its phosphatase CheZ [266]. Chemical and physical stimuli are detected by five chemoreceptors: Tar, Tap, Tsr, Trg and Aer [266]. Most of the chemotaxis genes are organized into two polycistronic transcriptional units (operons) [148]: *meche*, encoding Tar, Tap, CheR, CheB, CheY, and CheZ, and *mocha*, encoding CheA, CheW, and the flagellar motor proteins MotA and MotB. Tsr, Trg and Aer are encoded elsewhere on the genome.

To understand the differences in the chemotactic behavior of the selected ECOR strains at the genetic level, we sequenced the major chemotaxis operon *meche* of each of these strains (Figure B.2A). The total length of the *meche* operon is conserved in 16 of the 18 strains; the other two strains (ECOR51 and ECOR62), which are both from group B2, have a truncation of the gene of the chemoreceptor Tap. The truncated region is flanked on both sides by the same sequence GAATCAGG, likely a site of homologous recombination that led to the deletion of part of the gene. The truncated *tap* gene nevertheless is in-frame and may produce a 75-amino acid peptide (42 amino acids from the start and 33 amino acids from the end of the gene; Figure B.2A), in which the first transmembrane domain is preserved.

We observed also a truncation of gene product due to a mutation in the start codon. The *cheB* gene of ECOR62 had a reading frame starting at the 43rd base pair of the *cheB* gene of the rest of the strains. This truncation might affect the CheB expression level [208]. The truncated amino acids 4 – 42 belong to the regulatory (REC) domain of CheB (see Appendix C). However, according to a domain prediction using the SMART database [141,217], the two-domain structure of CheB is preserved in ECOR62.

We analyzed the polymorphic sites at the nucleotide and amino acid sequences of the ECOR *meche* genes, compared with the *meche* genes from the model strain for chemotaxis, *E. coli* K12. For the selected subset of strains, 4.1% of polymorphic sites exist at the nucleotide sequence of the

whole *meche* operon, ranging from 2.8% (for *cheY*) to 5.8% (for *cheB*) per gene (Figure B.2B). At the amino acid level, polymorphic sites make up 1.9% of all sequences, ranging from 0.5% (for CheZ) to 2.8% (for Tap) per protein (Figure B.2B, polymorphic sites at the amino acid level are indicated as "informative sites"). The group D strain ECOR48 stands out with the highest number of polymorphisms at the nucleotide level but not at the amino acid level. The high rate of third-base (non-coding) substitutions, *i.e.* different codon usage, might indicate the occurrence of horizontal gene transfer of the *meche* operon in the evolutionary past of this strain [174]. However, ECOR48 also shows high occurrence of substitutions on nucleotide level and amino acid level in other parts of the genome (in the fragments of housekeeping genes discussed below).

To compare the evolution of the chemotaxis genes with the evolution of the rest of the genome, we performed an identical analysis of nucleotide and amino acid polymorphisms in concatenated fragments of seven housekeeping genes of the selected ECOR strains (the sequences of *recA*, *purA*, *mdh*, *icd*, *gyrB*, *fumC*, and *adk* gene fragments for 492 *E. coli* isolates are publicly available in the MSLT database [275]) (Figure B.2B). These genes are distributed over the *E. coli* chromosome, and a previous study showed that their concatenated sequences fell largely into four phylogenetic clades in good agreement with the MLEE groups A, B1, B2 and D [275]. Polymorphisms at the nucleotide level for the selected subset of ECOR strains were found to be 4.2% (from 2.3% to 7.0% per gene fragment), very similar to the nucleotide polymorphism level of the *meche* operon (4.1%). At the amino acid level, the percentage of polymorphic sites for the housekeeping genes is lower (0.7%, ranging from 0.6 to 1.7% per fragment); however, a larger sequence sample might be needed to confirm the significance of this difference.

We used the sequence information for the *meche* operon to construct a phylogenetic tree (see Materials and methods) of the selected subset of ECOR strains (Figure B.3). Strains are clustered in a very good agreement with the MLEE phylogenetic groups (color-coded on Figure B.3). The laboratory strain *E. coli* K12 and a closely related species, *Salmonella typhimurium* LT2, are included for comparison. As expected, *E. coli* K12 clusters with the group A strains. However, we do not observe a

pronounced correlation between the chemotactic spreading and random motility characteristics with the sequence divergence of the *meche* operon, as illustrated in Figure B.3.

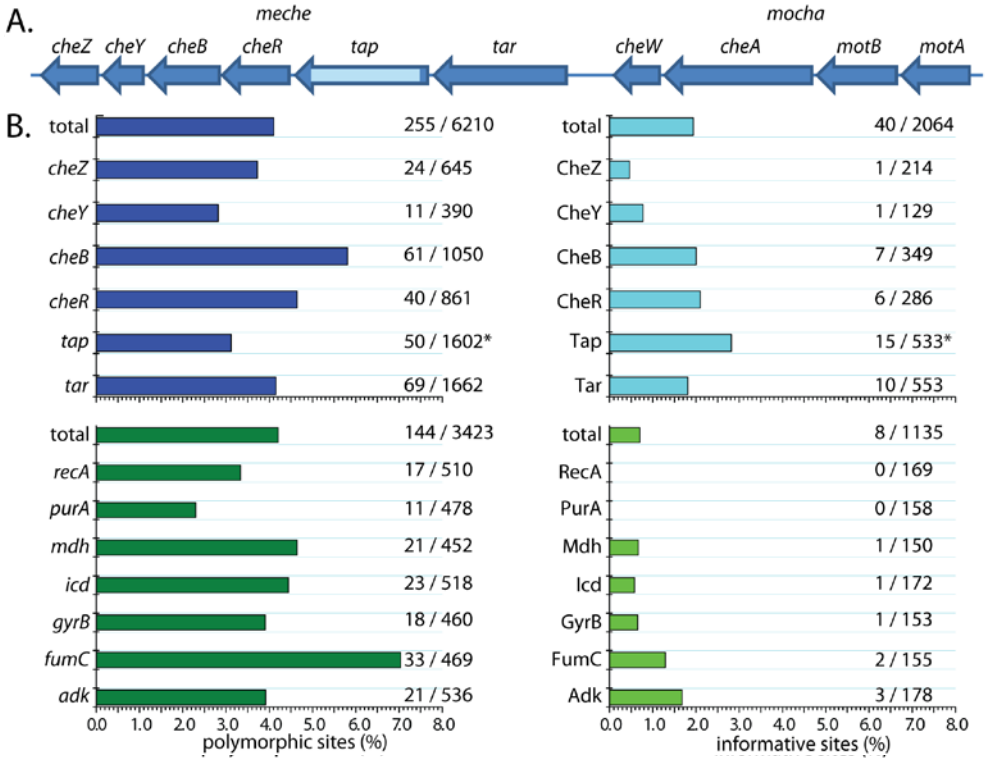

Figure B.2. Genetic diversity of chemotaxis genes in the selected ECOR strains. **(A)** Structure of the *meche* operon and the adjacent *mocha* operon of *E. coli*. The light blue box indicates the truncation at the *tap* gene of ECOR51 and ECOR62. **(B)** Histograms of the polymorphic sites for *meche* operon genes (top left) and housekeeping genes from MSTL database [275] (bottom left). Histograms of the polymorphic sites at the respective amino acid sequences ("informative sites"), are shown for the proteins expressed from the *meche* operon (top right) and housekeeping genes (bottom right). The number of substitutions over the total number of base pairs or amino acids respectively is shown. The first rows (labeled as "total") show the total number of polymorphic sites at all the sequences.

Appendix B

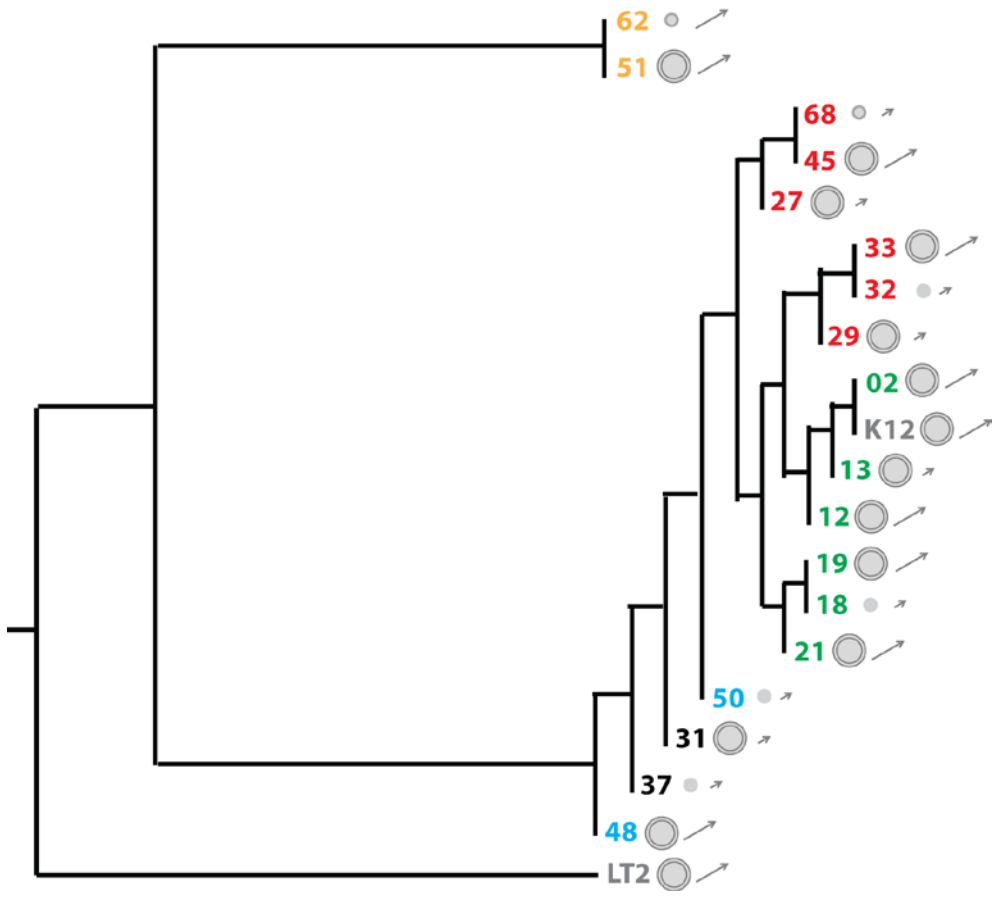

Figure B.3. Phylogenetic tree of the selected subset of ECOR strains, including the laboratory strain *E. coli* K12 and its closely related species *S. typhimurium* LT2. The tree was constructed using the concatenated sequences of *meche* genes of the selected strains and an algorithm, implemented into the BioEdit software (Fitch-Margoliash and least squares methods with evolutionary clock–based algorithm). The MLEE phylogenetic groups are color coded as in Tables B.1 and B.2. Spreading and swimming motility are indicated next to each strain as following: large circles – a strain forming large rings, small circles – a strain forming small but distinct rings, one circle – a strain that do not form distinct rings but spread diffusely; a large arrow – a strain that swims fast (swimming speed ≥10 μm/s), a small arrow – a strain that swims slow (swimming speed <10 μm/s) (see Table B.2).

B.5. Discussion

We explored the natural variation in the chemotactic behavior of *E. coli*, using a subset of the strains from the ECOR collection that represents much of the contemporary genetic diversity of the species, and covers strains from different locations and host origins. The chemotactic performance differs largely between the strains, even between those that are close phylogenetic relatives, and have comparable swimming speeds. Our observations suggest that after the divergence from their common ancestor and a major historical bottleneck ~30 million years ago that removed most of the extant diversity [57,275], the chemotaxis system of *E. coli* strains has evolved, likely improving the fitness (chemotactic performance) of the bacteria in the various environmental niches that different strains inhabit [70,256,273].

We analyzed sequences of the major genes of the chemotactic signaling system, which are part of the *meche* operon: the genes of the chemoreceptors Tar and Tap, the adaptation enzymes CheR and CheB, the response regulator CheY, and its phosphataze CheZ. Overall, the nucleotide substitution frequencies were found to be similar to those reported for housekeeping genes [275]. The adaptive evolution in prokaryotes is driven by mutation and recombination [107], the latter underlying the process of horizontal gene transfer [38]. We observed examples for sequence variations in the chemotaxis genes that likely result from mutation or recombination processes. For example, a mutation in the start codon of the *cheB* gene of ECOR62 creates a truncated open reading frame. Mutations in a close proximity to the start codon might affect the translational expression [208,209], therefore the expression levels of the methylesterase CheB might be altered in ECOR62, which can affect the methylation / demethylation balance and, in turn, the modification states of the chemoreceptors [215].

Recombination-based changes are the likely cause of the truncation of the *tap* gene in ECOR51 and ECOR62, and the high level of third-base substitutions (*i.e.* different codon usage) in the *meche* operon of ECOR48. Truncations of the genes of Tap and Trg chemoreceptors were previously reported to be prevalent in uropathogenic *E. coli* strains [133]; ECOR62 is also

an isolate from a pyelonephritis patient [189]. The truncation has occurred between the same flanking sequences, GAATCAGG, in the *tap* gene. The *tap* reading frame is preserved, as well as the first transmembrane domain, suggesting that the resulting 75-amino acid peptide might integrate into the membrane. The occurrence of clustered third-base substitutions might also indicate genetic changes due to recombination [174]. We observed multiple third-base substitutions in the strain ECOR48. This strain also contains a F1-like plasmid [159], which could take part in a conjugative horizontal gene transfer.

Our analyses of the chemotactic behavior and the genetic variation of the chemotaxis operon *meche* of *E. coli* provide a basis for future physiological studies that could assess which quantitative features of chemotaxis have undergone changes in the recent evolution of this species. We have used *in vivo* fluorescence resonance energy transfer (FRET) [242] to perform a detailed comparison between the quantitative features of the chemotactic signaling of *E. coli* K12 and the closely related species *S. typhimurium* LT2, and we have identified differences in the transfer functions of chemotactic signaling [260] that affect the sensitivity to chemoeffectors, the receptor-receptor cooperativity, and the adaptation kinetics (Chapter 3). The chemotactic transfer functions of *S. typhimurium* could serve as an outgroup for a similar comparison of the chemotactic transfer functions of the selected *E. coli* strains (we have engineered genetically a subset of these strains for future FRET experiments; see Materials and methods and Chapter 6). The fitness (chemotactic performance) of the compared strains can be evaluated quantitatively in defined gradients created in microfluidic devices [8]. Such future studies could reveal design principles underlying the evolution of the bacterial chemotaxis system, and improve our understanding for the design of signaling circuits in nature.

B.6. Materials and methods

Bacterial strains

The ECOR collection (Table B.1) is described in reference [189].

Some of the ECOR strains (listed below) were engineered genetically for future FRET experiments in the following manner. In-frame chromosomal gene deletions of *cheY* and *cheZ* genes were created using an allele replacement procedure, based on a modification of Datsenko and Wanner's method [62], that does not leave a scar. It is based on an insertion cassette that contains the lethal *ccdB* gene under the control of a L-rhamnose-inducible promoter. This cassette can be removed later by positive selection on rhamnose-minimal plates [288]. Then the plasmid pVS88 [240] used for expression of the FRET fusion proteins CheY-YFP and CheZ-CFP was transformed in the selected strains.

ECOR strains 13, 18, 19, 21, 27, 29, 32, 33, 48, 51, 62 and 68 were engineered for FRET in the way described above. ECOR strains 2, 12, 45 and 50 were engineered in a similar manner, but the *ccdB* cassettes were not removed because of problems with false positive colonies that have occurred with the rhamnose positive selection strategy. The strains ECOR31 and 37 are resistant to ampicillin and kanamycin and chloramphenicol and kanamycin respectively, and thus these strains have not been engineered for FRET experiments (the plasmids used in the knockout strategy have ampicillin and kanamycin resistances, and pVS88 has an amplicillin resistance, thus the native resistances of ECOR31 and 37 should be removed prior to knock out or transformation procedures).

All strains engineered for FRET should be tested to determine the optimal levels of induction of the FRET fluorescent fusions. The pilot experiments with ECOR13, 18, 29, and 68 showed that the same amount of inducer (IPTG) led to ~two-fold lower level of fluorescence for both yellow and cyan channels. The fluorescence levels in ECOR51 were similar to that of *E. coli* HCB33 (wild type for chemotaxis). The attachment of the strains to the coverslips in the FRET experiments should also be optimized (e.g. ECOR18 do not stick well to polylysine-coated coverslips).

Appendix B

Soft-agar plate experiments

The bacterial cultures were grown overnight to saturation in tryptone broth (TB, 1% Bacto tryptone, 0.5% sodium chloride, pH 7.0). The soft-agar plates were prepared using 25 ml TB per plate, solidified with 0.26% Bacto agar. The plates were left to solidify overnight under a humidity chamber. 5 µl of the tested cultures were inoculated in the plates. Six strains were tested per plate, and two independent replicates were performed for each strain. The plates were incubated at 30° C for 4 h, and imaged using MultiDoc-It™ digital imaging system, equipped with Olympus camera. After an additional 3.5 h of incubation, the plates were imaged again.

Plate reader experiments

The growth in TB at 33.5° C was evaluated for the strains of interest by measuring the optical density of the cultures at 600 nm (OD_{600}) in a multilabel plate reader (Perkin Elmer 2030). The temperature of the plate reader was set to 33.5° C. The saturated overnight cultures were diluted 100 times in TB. 200 µl of the diluted cultures were placed in the wells of a 96-well plate (OptiPlate-96). The OD_{600} was measured and recorded at 35 min intervals. After every third measurement, 11 µl of water was dispensed in each well to compensate for the evaporation of water and maintain the volume in each well constant. One of the rows in the 96-well plate was filled with a sterile TB and the averaged OD_{600} measured in this row was used for a background correction. At least four wells were allocated for each strain and the averaged results were used to plot a growth curve for each strain. The doubling times of the strains were calculated by fitting to the log phase of the growth curves an exponential function $OD_{600} = OD_{600}^{0} 2^{\tau}$, where OD_{600}^{0} is the OD_{600} in the beginning of the log phase and τ is the doubling time of the strain.

Swimming motility experiments

The overnight cultures were diluted 100x in TB and grown at 33.5° C on a rotary shaker, until the cultures reached mid-exponential phase (OD_{600} ~ 0.5). Then the bacteria were harvested by centrifugation (5 min, 5000 rpm), washed twice and carefully resuspended (by shaking the tube, without pipetting) in motility buffer (10 mM potassium phosphate buffer pH 7.0, 0.1 mM EDTA, 1 μM *L*-methionine, 10 mM lactic acid, pH 7.0). The cultures were imaged immediately or stored at 4° C up to 5 h prior to the experiment.

The swimming behavior of the bacteria was observed on a Nikon Eclipse inverted microscope using a 20x phase-contrast objective and an additional 1.5x magnification. The tested bacterial strain, diluted 3 - 5x in motility buffer was placed in tunnel slides, in which the distance between the glass and the cover slip was ~1.5 mm. The bacteria were imaged at mid-depth, close to the liquid-air border of the tunnel slide to ensure that oxygen was available to the bacteria. The temperature during the microscopy observation was maintained at 33.5° C using a temperature-control plastic chamber covering the microscope.

A Roper Scientific (Photometrics CoolSNAP HQ) camera was used to record movies with 7.6 frames per second. Movies with a higher frame rate (30 frames per second) were also recorded for some of the strains using a smaller field of view. The bacterial trajectories were extracted using BacTrack cell tracking software [9]. First, each frame was subtracted from the following one to remove the background; then, the bacteria in each frame were located as peaks in a monochrome intensity field; finally, the bacteria were tracked between the frames using the particle tracking algorithm "Conservative search radius" (also called "Roman1") implemented in the BacTrack software. This algorithm utilizes a user-defined hard cutoff radius that corresponds to the maximum distance that an object (bacterium) can move between successive frames. The objects in a frame *n* are called "parent particles" and the objects in the consecutive frame, *n+1*, are called "children particles". The basic rule is that each child comes from one parent and each parent begets one child. The algorithm does the following:

Appendix B

1. locates parent and children particles in frames *n* and *n+1*
2. accepts an user-defined cutoff radius *R*
3. computes the distances for all pairs of parents and children; throw out all pairs whose distance > *R*
4. considers for one parent particle at a time whether the parent has
   (a) no children (no child particle within *R*): the trajectory ends
   (b) a single child (only one child particle within *R*); if the child has
      - a single parent, the child-parent pair is considered a match (the child and the parent are the same particle)
      - >1 parent: the trajectory ends
   (c) >1 child (>1 child particle within *R*): the trajectory ends
5. the children that have no parent start new trajectories
6. the process is repeated from step 2 for the next pair of frames

The post-processing of trajectories was performed in MATLAB to yield the 2D population-averaged run velocity (the MATLAB code that was used has been published in the B.Sc. thesis of M. D. Sekora, MIT, 2005).

The tracking algorithm has worked only for a subset (~1/4 of the total number) of the movies of swimming ECOR strains. The main problems were the density of the cells (too high or too low) and drift of the liquid in the tunnel slides. Fast / slow categories (Table B.1) are thus approximate, strains that seemed (by eye) to swim similar to strains with swimming speed <10 μm/s were labeled as "slow", and strains that seemed to swim similar to strains with swimming speed ≥10 μm/s were labeled as "fast".

Sequencing and genetic analysis

Sequencing primers for the *meche* region were designed using the Primer 3 program and the genomic sequence of the *E. coli* K12 substrain MG1655 (NCBI accession number: NC_000913.2). PCR products of length 700-800 base pairs were obtained for the whole length of the *meche* operon, including its flanking regions, and the samples were sent for sequencing to ServiceXS. Results were analyzed using BioEdit and Mutalin programs. Multiple sequence alignments were obtained using CLUSTALW algorithm. ExPASy translate tool was used obtain the amino acid sequences.

The concatenated sequences of the genes of *meche* operon from the selected subset of ECOR strains, *E.* coli K12 substrain MG1655, and *S. typhimurium* LT2 were used to construct a neighbor-joining tree using KITSCH© algorithm by Joseph Felsenstein in Bioedit, which estimates phylogenies from distance matrix data, assuming that the distances are equal to the sums of branch lengths between the strains. It assumes an evolutionary clock and uses the Fitch-Margoliash criterion.

B.7. Acknowledgements

Michel de Vries provided help with the sequencing of the chemotaxis operon of ECOR strains.

Appendix B