Supplemental files

Figure S1. Diversity of carbon sources used by 88 *E. coli* and *Escherichia* clade strains. Twelve carbon sources enabled the growth of all 88 strains (common substrates), while 59 were able to grow on only a fraction of them (selective substrates). The strains are ordered by number of substrates they can catabolize. The carbon sources are ordered by number of strains able to grow on them.

Figure S1

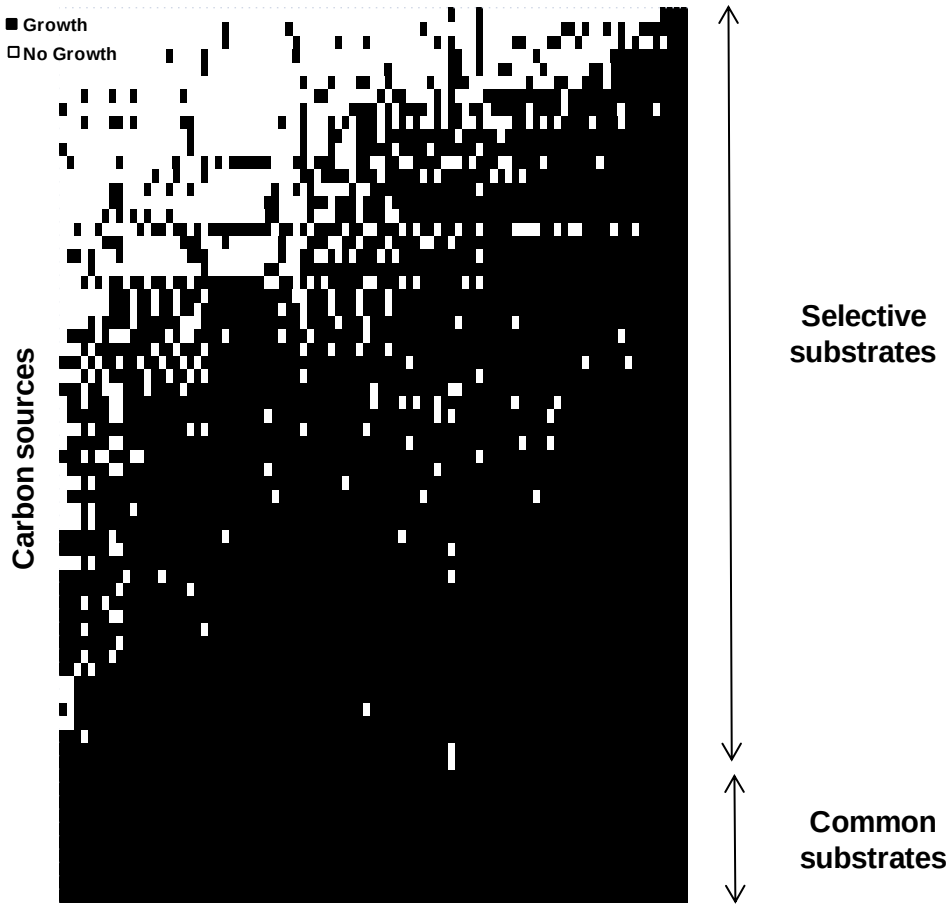
Growth
No Growth
Carbon sources
Selective substrates
Common substrates

Table S1. Phylo-group distribution of *E. coli* and *Escherichia* clade water-collection and selection of microcosm subsample

Table S2. List of primers used in the study for PCR O-typing. All *E. coli* strains were tested for 28 O-types by using the molecular O-group determination method [34]

Table S3. Physicochemical characteristics of water from vulnerability proxies

Table S4. Diversity 88 *E. coli* and *Escherichia* clade strains from the microcosm experiment (A) waterbodies samples, (B) control samples

Table S5: Prevalence of virulence genes according to survival abilities of 58 *E. coli* and *Escherichia* clade strains isolated from the four waterbodies

Table S6. Predation by *Dictyostelium discoideum* of 88 *E. coli* and *Escherichia* clade strains according to origin and survival abilities

Table S1. Phylo-group distribution of *E. coli* and *Escherichia* clade water-collection and selection of microcosm subsample

Water profile (N/n [a])		Distribution of phylo-groups of *E. coli* collection											
		Total library (352)						Microcosm experiment (88)					
		A	B1	B2	D	Clade III	Clade V	A	B1	B2	D	Clade III	Clade V
1 (45/15)		3	39	1	2	0	0	1	13	0	1	0	0
2 (39/13**)**		10	6	4	8	1	10	1	2	2	3	0	5
3 (46/15)		22	20	1	3	0	0	7	7	0	1	0	0
4 (34/15)		20	4	2	5	0	3	9	1	1	2	0	2
Control (N/n [a])		**A**	**B1**	**B2**	**D**	**Clade III**	**Clade V**	**A**	**B1**	**B2**	**D**	**Clade III**	**Clade V**
Human (50 /15)	Medical center wastewater	28 + (2 [c])	4	15	1	0	0	7 + (2 [c])	1	5	0	0	0
Bovine (138 /15)	Cowpats [b]	7	118	0	13	0	0	1	13	0	1	0	0

[a] N: total number of collected strains, n number of randomly selected strains based on phylo-group distribution, [b] mix from dry and fresh cowpats, [c] *Escherichia* belonging to phylo-group A by the classical triplex Clermont method [4] but identified as phylo-group C [31,32].

Table S2. List of primers used in the study for PCR O-Typing. All *E. coli* strains were tested for 28 O-types by using the PCR-based O group determination method [34].

Primer designation	Primer sequence	Target	Size of PCR product (bp)	Reference
gndbis.f	5'-ATACCGACGACGCCGATCTG-3'			[34]
rfbO1.r	5'-CCAGAAATACACTTGGAGAC-3'	*rfbO1*	189	[34]
rfbO2a.r	5'-GTGACTATTTCGTTACAAGC-3'	*rfbO2*	274	[34]
rfbO4.r	5'-AGGGGCCATTTGACCCACTC-3'	*rfbO4*	193	[34]
rfbO6a.r	5'-AAATGAGCGCCCACCATTAC-3'	*rfbO6*	584	[34]
rfbO7.r	5'-CGAAGATCATCCACGATCCG-3'	*rfbO7*	722	[34]
rfbO8a.r	5'-GAACAATATTGTAAGGTCGCC-3'	*rfbO8*	227	[33]
rfbO12.r	5'-GTGTCAAATGCCTGTCACCG-3'	*rfbO12*	239	[34]
rfbO15.r	5'-GTTTACGTTCCCACCTTATG-3'	*rfbO15*	486	[34]
rfbO16.r	5'-GGATCATTTATGCTGGTACG-3'	*rfbO16*	450	[34]
rfbO18.r	5'-GAAGATGGCTATAATGGTTG-3'	*rfbO18*	360	[34]
rfbO25a.r	5'-GAGATCCAAAAACAGTTTGTG-3'	*rfbO25*	313	[34]
rfbO25b.r	5'-TGCTATTCATTATGCGCAGC-3'	*rfbO25*	300	[33]
rfbO26.r	5'-GTATGAGCAAAATGGTGAGC-3'	*rfbO26*	329	[33]
rfbO40.r	5'- CAGGAAAGCCTCACTATTGG-3'	*rfbO40*	625	This study
rfbO45b.r	5'-TGCGAGTAGACTATCTCAAG-3'	*rfbO45*	436	[33]
rfbO55r	5'-AATTTAGGTCCGGCAGCAAG-3'	*rfbO55*	675	[33]
rfbO75.r	5'-GTAATAATGCTTGCGAAACC-3'	*rfbO75*	419	[34]
rfbO78.r	5'-GCACTGCCATTGGTATTTACG-3'	*rfbO78*	464	[33]
rfbO81.r	5'-GAGCAGTATATATTACTGGTG-3'	*rfbO81*	383	[33]
rfbO86.r	5'-CGTTGTTAATAATTCTGAATGCG-3'	*rfbO86*	361	[33]
rfbO88.r	5'-AAGGAAAAACGCTGGGAGAG-3'	*rfbO88*	494	[33]
rfbO102.r	5'-TACCCATGATGGTACTGGTG-3'	*rfbO102*	480	[33]
rfbO103.r	5'-GAACTTGGATGGAAAGCCTG-3'	*rfbO103*	242	[33]
rfbO104.r	5'-TGGCTTAGGATACTTGCAGC-3'	*rfbO104*	410	[33]
rfbO111.r	5'-TTGAGTTCTGAGTGGGAAGG-3'	*rfbO111*	522	[33]
rfbO128.r	5'-AACTCGGAGAGTTCCCTATG-3'	*rfbO128*	319	[33]
rfbO150.r	5'-TAACGCTAGTGGCAGCAATG-3'	*rfbO150*	602	[33]
rfbO157.r	5'-TACGACAGAGAGTGTCTGAG-3'	*rfbO157*	672	[34]

Table S3. Physicochemical characteristics of water from vulnerability proxies

Sampling site		T°C	pH	Conductivity µS.cm^{-1}	SiO_2	Mg^{2+}	Ca^{2+}	K^+	Na^+	Fe^{3+}	NH_4^+	NO_3^-	NO_2^-	HCO_3^-	PO_4^{3-}	SO_4^{2-}	Cl^-	DOM
					mg.L^{-1}													
Creek water	Low flow period	14	8.13	345	12.3	3.55	47.5	1.8	13.45	0	0	12.9	0	112	0.14	29.7	22.3	0.1
	High flow period	13.5	7.57	557	7.14	3.25	36.05	4.45	9.45	0.1	0.1	13.75	0	68.3	0.21	33.3	20.6	nd
River		9.8	8.24	531	13.9	5.5	97	2.1	10.7	10	0.05	22.9	0.03	0.5	0.153	14.6	164	<0.5
Estuary		8.9	7.78	7500	7.1	165.8	136	52.2	1378	nd	0.05	26	0.31	0.5	0.307	377	2520	3.16

DOM: Dissolved Organic Matter, nd: not determined

Table S4. Diversity of 88 *E. coli* and *Escherichia* clade strains from the microcosm experiment (A) waterbodies samples, (B) control samples

(A)

Sample site	*E. coli* phylo-group	*uidA* allele	O-type	N	Epidemiological type	Survival group *
Type of water contamination 1	A_1	*uidA*101	NT	1	ET24	S1
	B1	*uidA*110	NT	1	ET50	S2
		*uidA*103	NT	1	ET49	S2
		*uidA*90	**O150**	1	ET47	S1
		*uidA*90	**O150**	1	ET47	S2
		*uidA*90	**O150**	1	ET48	S2
		*uidA*90	**O15**	1	ET46	S1
		*uidA*90	**O15**	2	ET46	* S3 [(a)]
		*uidA*40	**O40**	1	ET38	S2
		*uidA*2	**O40**	1	ET27	S2
		*uidA*5	NT	1	ET32	S2
		*uidA*2	NT	1	ET28	S1
		uidA2	NT	1	ET28	S2
	D	*uidA*2	NT	1	ET58	S2
Type of water contamination 2	A_0	*nd*	NT	1	ET15	S4
	B1	*uidA*30	**O7**	1	ET36	S2
		*uidA*30	NT	1	ET37	S4
	B2	*uidA*6	NT	1	ET54	S3
		*uidA*6	NT	1	ET53	S4
	D	*uidA*91	NT	1	ET65	S2
		*uidA*65	NT	1	ET63	S4
		*uidA*13	**O86**	1	ET59	S3
	Escherichia clade V	*uidA*111	NT	2	ET66	* S2 [(a, b)]
		*uidA*111	NT	1	ET66	S3
		*uidA*111	NT	1	ET66	S4
		*uidA*111	NT	1	ET68	S4
Type of water contamination 3	A_0	*uidA*77	NT	1	ET14	S1
		*uidA*40	NT	1	ET12	S1
		*uidA*22	NT	1	ET10	S4
		*uidA*20	NT	1	ET8	S2
	A_1	*uidA*77	**O6a**	1	ET22	S4
		*uidA*2	**O12**	1	ET18	S2
	B1	*uidA*55	NT	1	ET41	S3
		*uidA*55	NT	1	ET40	S4
		*uidA*55	NT	1	ET42	S4
		*uidA*50	NT	1	ET39	S3
		*uidA*23	NT	1	ET35	S2
		*uidA*23	NT	1	ET35	S4
		*uidA*5	NT	1	ET31	S2
		*uidA*2	NT	1	ET26	S3
	D	*uidA*32	NT	1	ET61	S3
Type of water contamination 4	A_0	*uidA*2	NT	1	ET2	S3
		*uidA*2	**O6a**	1	ET1	S4
		*uidA*16	NT	1	ET6	S3
		*uidA*20	**O18**	1	ET7	S4
		*uidA*40	**O6a**	1	ET11	S2
	A_1	*uidA*2	NT	1	ET17	S2
		*uidA*2	**O6a**	1	ET16	S4
		*uidA*20	NT	1	ET20	S3
		*uidA*23	**O6a**	1	ET21	S2
	B1	*uidA*2	NT	1	ET25	S2
	B2	*uidA*80	NT	1	ET56	S3
	D	*uidA*42	**O18**	1	ET62	S3
		*uidA*14	**O1**	1	ET60	S2
	Escherichia clade V	*uidA*111	NT	1	ET67	S2
		*uidA*111	NT	1	ET67	S4

(B)

Sample site	*E. coli* phylo-group	*uidA* allele	O-type	N	Epidemiological type	Survival group
Medical center waste water	A_0	*uidA*40	**O18**	1	ET13	S4
		*uidA*20	**O102**	1	ET9	S4
		*uidA*4	NT	1	ET4	S3
		*uidA*3	NT	1	ET3	S2
	A_1	*uidA*77	NT	1	ET23	S1
		*uidA*77	NT	1	ET23	S4
		*uidA*2	NT	1	ET19	S1
	B1	*uidA*2	NT	1	ET30	S1
	B2	*uidA*6	NT	1	ET55	S1
		*uidA*1	**O4**	1	ET52	S2
		*uidA*1	**O6a**	3	ET51	* S4 [(a)]
	C	*uidA*4	**O6a**	1	ET57	S2
		*uidA*4	**O6a**	1	ET57	S3
Cowpats	A_0	*uidA*5	**O7**	1	ET5	S3
	B1	*uidA*5	**O7**	1	ET33	S3
		*uidA*5	**O7**	2	ET33	* S4 [(a)]
		*uidA*5	NT	1	ET34	S2
		*uidA*55	NT	1	ET43	S2
		*uidA*55	NT	2	ET43	* S3 [(a, b)]
		*uidA*55	NT	2	ET44	* S2 [(a, b)]
		*uidA*55	NT	2	ET45	* S2 [(b)]
		*uidA*55	NT	1	ET45	S3
		*uidA*2	NT	1	ET29	S2
	D	*uidA*89	NT	1	ET64	S2

A_0 and A_1 phylo-groups were distinguished as exhibiting (---) and (-+-) genotypes, respectively, as in [5]. C phylo-group was distinguished as in [32]. NT: O1, O2a, O4, O6a, O7, O8a, O12, O15, O16, O18, O25a, O25b, O26, O40, O45b, O55, O75, O78, O81, O86, O88, O102, O103, O104, O111, O128, O150 or O157 negative; N: number of strains. * Strains were differentiated on the basis of [(a)] phenotypic traits or [(b)] the antibiotic resistance profiles. nd: not determined.

Table S5: Prevalence of virulence genes according to survival abilities of 58 *E. coli* and *Escherichia* clade strains isolated from the four waterbodies.

Gene	Description of the encoded protein	Presence in survival group			
		S1	**S2**	**S3**	**S4**
		(N= 6)	**(N= 23)**	**(N= 14)**	**(N= 15)**
sfa/foc	S fimbrial adhesin D-E proteins/F1C fimbria proteins	-	-	2	1
iroN	Iron-related siderophore receptor	-	1	4	2
iutA	Ferric aerobactin receptor	-	1	3	1
papC	PapC protein (formation of digalactoside-binding Pap pili)	1	1	3	-
hlyC	Hemolysin C	-	-	-	1
cnf1	Cytotoxic necrotizing factor 1	-	-	-	-
fyuA	Yersinabactin receptor	2	3	5	4
kspE	Capsular polysaccharide export system inner membrane protein	3	14	9	6
ompT	Outer membrane protease	4	14	6	7
afaD	Afimbrial adhesin D invasin	-	-	1	-

N: total number of strains; -: not detected;

Table S6: Predation by *Dictyostelium discoideum* of 88 *E. coli* and *Escherichia* clade strains according to origin and survival abilities

	***E. coli* origin**													
	Waterbodies (58 strains)						**Control (30 strains)**							
							Human				**Bovine**			
	N (%) (n=58)	**Survival groups**				**N (%)** (n=30)	**Survival groups**							
		S1 (n=6)	**S2** (n=23)	**S3** (n=14)	**S4** (n=15)		**S1** (n=4)	**S2** (n=3)	**S3** (n=2)	**S4** (n=6)	**S1** (n=0)	**S2** (n=8)	**S3** (n=5)	**S4** (n=2)
Resistance to predation	39 (67.2)	3	14	12	10	14 (46.6)	3	3	2	3	0	2	1	0

N: total number of strains