Metabolic Engineering Communications 5 (2017) 78–83

Contents lists available at ScienceDirect

Metabolic Engineering Communications

journal homepage: www.elsevier.com/locate/mec

Genome sequence and analysis of *Escherichia coli* production strain LS5218

Jacqueline M. Rand[a], Gina C. Gordon[a,b], Christopher R. Mehrer[a], Brian F. Pfleger[a,b,*]

[a] Department of Chemical and Biological Engineering, University of Wisconsin, Madison, United States
[b] Microbiology Doctoral Training Program, University of Wisconsin, Madison, United States

ARTICLE INFO

Keywords:
E. coli K-12
Genome sequence
Polyhydroxyalkanoate
Metabolic engineering

ABSTRACT

Escherichia coli strain LS5218 is a useful host for the production of fatty acid derived products, but the genetics underlying this utility have not been fully investigated. Here, we report the genome sequence of LS5218 and a list of large mutations and single nucleotide permutations (SNPs) relative to *E. coli* K-12 strain MG1655. We discuss how genetic differences may affect the physiological differences between LS5218 and MG1655. We find that LS5218 is more closely related to *E. coli* strain NCM3722 and suspect that small genetic differences between K-12 derived strains may have a significant impact on metabolic engineering efforts.

1. Introduction

Escherichia coli strain LS5218 is frequently studied for the production of polyhydroxyalkanoates (PHAs) from mixtures of sugars and organics acids (Agnew et al., 2012; Nduko et al., 2012; Salamanca-cardona et al., 2014). LS5218 is selected because of two commonly cited differences from other *E. coli* strains – mutations in *fadR* (*fadR601)* and *atoC* (*atoC(c)*). The *fadR601* disrupts expression of FadR thereby deregulating the *fad* genes that encode enzymes responsible for β-oxidation (Fujita et al., 2007). AtoC is an activator of the *atoDAEB* operon, encoding enzymes required for catabolism of acetoacetate and other short-chain organic acids (Lioliou et al., 2005; Theodorou et al., 2011). The *atoC(c)* mutation alters the regulator and causes constitutive expression and upregulation of the *atoDAEB* operon (Jenkins and Nunn, 1987; Matta et al., 2007). The mutations in *E. coli* LS5218 allow for increased uptake and utilization of a wider array of fatty acid chain-lengths and make it well-adapted for the engineering of short chain length-co-medium chain length (SCL-co-MCL) copolymers and poly(3-hydroxybutyrate-co-3-hydroxyvalerate) [P(3HB-co-3HV)] (Rhie and Dennis, 1995; Tappel et al., 2012). Despite its widespread use in PHA production studies, the genome sequence of *E. coli* LS5218 has not been made publicly available. This is in part due to the common assumption that it is a close relative of the sequenced *E. coli* K-12 strain MG1655.

While a variety of *E. coli* strains are widely used by researchers, the history of their isolation is not as widely known. The original *E. coli* K-12 was isolated in 1922 and deposited in the Stanford University strain collection in 1925 (Neidhardt et al., 1996). The two main wild-type K-12 strains, WG1 from J. Lederberg and EMG2 from Clowes and Hayes, are subcultures of the Stanford K-12 strain. The published derivation of *E. coli* LS5218 involved a two-step screening of spontaneous mutants on selective media (Fig. 1**A**) (Spratt et al., 1981). Strain RS3010 was a spontaneous mutant of the Lederberg *E. coli* K-12 strain selected for growth on decanoate, to isolate a mutant with upregulated β-oxidation gene expression (Simons et al., 1980). Strain LS5218 was generated as a spontaneous mutant of RS3010 selected for on butyrate in order to isolate a strain with the ability to metabolize SCL fatty acids (Spratt et al., 1981). *E. coli* MG1655 was derived from an original K-12 isolate from the Lederberg lab through a two-step process designed to cure out the bacteriophage lambda (UV radiation and blood agar selection) and the F plasmid (acridine orange) (Blattner et al., 1997a). *E. coli* MG1655 and *E. coli* LS5218 appear to be derived from the same *E. coli* K-12 isolate (the Lederberg K-12 strain), but differences in their derivation histories convinced us to sequence *E. coli* LS5218 to know the exact genetic background of this production strain. Here, we report the genome sequence of *E. coli* LS5218 and an analysis of its content relative to *E. coli* MG1655 and a closer relative *E. coli* NCM3722.

2. Results and discussion

E. coli LS5218 genomic DNA was sequenced using paired end reads on a HiSeq. 2500 System, then assembled into 121 contigs using SPAdes (Bankevich et al., 2012). The draft genome was deposited in GenBank (GCA_002007165.1) and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) automatically assigned annotations. Using the annotated protein features for LS5218, we generated a phylogenetic tree comparing LS5218 with 21 completely sequenced *E. coli* K-12 derivatives using the Bacterial Pan Genome Analysis pipeline (BPGA) (Chaudhari et al., 2016). The pan genome analysis compiled a set of

* Corresponding author at: Department of Chemical and Biological Engineering, University of Wisconsin, Madison, United States.
E-mail address: pfleger@engr.wisc.edu (B.F. Pfleger).

http://dx.doi.org/10.1016/j.meteno.2017.10.001
Received 17 July 2017; Received in revised form 29 September 2017; Accepted 31 October 2017
Available online 02 November 2017

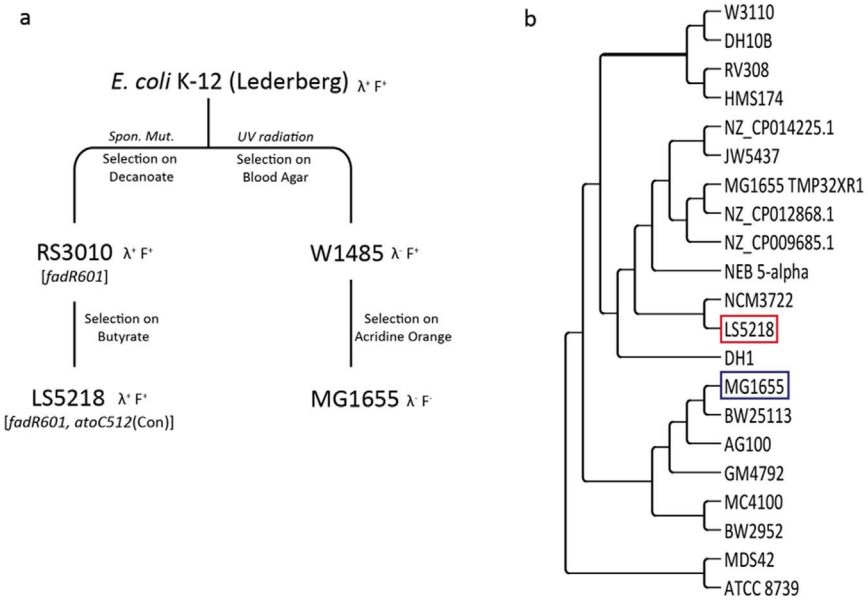

Fig. 1. Comparison of LS5218 and MG1655. A. Diagram of the published derivation paths for LS5218 and MG1655. B. Pan-genome phylogenetic tree for *E. coli* K-12 strains. Strains listed with published name, or accession number if a published name was not listed. Spon. Mut., spontaneous mutation.

core genes common in all strains, accessory genes present in at least two or more strains, and unique genes only found in a single strain. The algorithm used this information to perform phylogenetic clustering of the *E. coli* K-12 derivative strains based on their variable gene content (Fig. 1**B**). The phylogenetic clustering shows that LS5218 is more closely related to the newly sequenced strain *E. coli* NCM3722 than it is to MG1655.

Next, we compared the newly assembled LS5218 genome with the *E. coli* MG1655 reference genome to evaluate the genetic relationship between the two strains. We used the Mauve genome alignment software (Darling et al., 2004, 2010) to align the genome contigs for LS5218 against MG1655 and found 17 large differences in the form of gene insertions, deletions and genome rearrangements between the two strains (Table 1). We also performed single nucleotide polymorphism (SNP) and indel analysis with FreeBayes (Garrison and Marth, 2012), and found 74 small differences listed in Table 2. Through this analysis, we confirmed the presence of mutations in *fadR* and *atoC*.

The L55Q mutation in *fadR* replaces a hydrophobic leucine with a hydrophilic glutamine within the DNA binding domain. This change likely affects the interaction of *fadR* with the DNA backbone (van Aalten et al., 2000; Xu et al., 2001). The *atoC* mutation, I129S, is responsible for conferring constitutive expression of the *ato* operon, however the mechanism of this action remains unknown. Beyond the expected mutations, the major insertions and deletions were concentrated around insertion elements whereas the small SNPs were distributed evenly throughout the genome. Coverage of the LS5218 sequence compared to MG1655 (Fig. 2) highlights the position of known insertion elements in MG1655 for comparison of the large and small differences along with the assembled contigs.

We found that LS5218 has numerous insertions, deletions, genomic arrangements, and SNPs as well as the presence of the F plasmid. The highlighted 17 large insertions and deletions as well as the 74 SNPs could affect gene expression beyond the anticipated changes in fatty acid degradation pathways. One of the primary differences between MG1655 and LS5218 is the *rph* mutation. It is known that MG1655 has a frameshift mutation in *rph* that also causes pyrimidine starvation due to polar effects on the downstream *pyrE* gene (Blattner et al., 1997b). MG1655 also has a known frameshift in *ilvG* that affects expression of a branched-chain amino acid biosynthesis operon (Lawther et al., 1982), which is not present in LS5218. The fact that LS5218 does not have these mutations may partially explain why we see increased growth rates in LS5218 when compared with MG1655 on MOPS minimal media (Fig. 3).

Table 1
Table of Large insertions and deletions between MG1655 and LS5218.

Location	MG1655	LS5218	Comments
257905-258680	IS1I	No insert	Intact *crl* gene in LS5218
279599-291070[b]	No insert	Deletion – recombination at insA elements	Deletion of 11 genes of cryptic prophage CP4–6
574587-575785[b]	insH1	No insert	Intact *nmpC* gene in LS5218
687850-689049[b]	insH1	No insert	IS5 upstream of *gltLJKL* operon in MG1655
807329[b]	No insert	λ phage	Wild type λ phage in LS5218
916878[a]	No insert	Insertion in *ybjD*	Premature stop codon
1299498–1300697	IS5U	No insert	upstream of *oppA*
1878573[b]	No insert	IS5	Disrupted *yeaP* gene in LS5218
1978505–1979294[b]	IS1	Tn1000	Insertions upstream of *flhDC*
2101742–2102945[b]	IS5	No insert	Intact *wbbL* gene in LS5218
2110297–2128593[b]	No insert	IS1 and 18 kb deletion	Deleted: *rfbA, rfbD, rfbB, galF, wcaM, wcaL, wcaK, wzxC, wcaJ, cpsG, cpsB, wcaI, fcl, gmd, wcaF*
2170165–2171620[b]	IS3	No insert	Intact *gatR* in LS5218
3130145[a]	IS5	3.5 kb insert	Inserted: fatty acyl-AMP ligase, short chain dehydrogenase, ACP binding site family protein
3365549–3366752[b]	IS5	No insert	Intact *yhcE* gene in LS5218
4480807[b]	No insert	IS1	Disrupted *yjgN* gene in LS5218
4498173–4499513[b]	IS2	No insert	Insertion in MG1655 between pseudogenes in KpLE2
F Plasmid[b]	No	Yes	

[a] Similar position but different from reported mutation in NCM3722 (Lyons et al., 2011).
[b] , Mutation also reported for NCM3722 (Lyons et al., 2011).

Table 2
Table of SNPs and indels between MG1655 and LS5218.

Location	Gene	Type	CDNA change	AA change
280113	insX	CDS	AAGCTG→GGCTA	Lys82fs
1101543	csgG	CDS	A→T	Lys48[b]
1330578	yciN	CDS	ΔG	Ile31fs
2173360	gatC	CDS	ΔGG	Val306fs
2210942	yehQ	CDS	T→G	[b]615Glu
2278174	yejG	CDS	Insert CTGCTGGT	Phe22fs
2665747	csiE	CDS	C→T	Gln105[b]
2867455	[a]rpoS	CDS	C→T	Gln33[b]
3130140	yghO	CDS	A→T	Lys2[b]
3473612	[a]rpsG	CDS	T→A	Leu157[b]
3560455	glpR	CDS	Insert C	His51fs
3662700	[a]mdtF	CDS	C→T	Gln763[b]
3815879	[a]rph	CDS	Insert G	Glu224fs
3951535	[a]ilvG	CDS	Insert AT	Gln327fs
290103	argF	CDS	T→A	Phe68Tyr
290174	argF	CDS	TACAGAAGCTTACC→AAGCCAAACTCACT	ValGln40GluAla
290192	argF	CDS	ATGGCAAG→GCGGTAAA	Asn36Ser
290221	argF	CDS	AC→GA	Gln28Lys
378700	[a]frmA	CDS	T→G	Val291Gly
579285	ybcV	CDS	A→G	Ile104Val
616676	entF	CDS	C→A	Asp840Glu
903248	[a]artP	CDS	C→A	Leu163Met
1169836	ldtC	CDS	T→C	Leu180Pro
1235101	fadR	CDS	T→A	Leu55Gln
1301992	[a]oppA	CDS	A→T	Asn271Tyr
1301999	[a]oppA	CDS	G→A	Ser273Asn
1302190	[a]oppA	CDS	A→G	Asn337Asp
1305442	[a]oppD	CDS	T→G	Val230Gly
1306736	[a]oppF	CDS	T→G	Ser325Ala
1337394	[a]acnA	CDS	A→G	Ser522Gly
1358859	[a]puuP	CDS	A→G	Tyr110Cys
1643679	[a]ydfU	CDS	T→A	Leu209Gln
1652331	[a]intQ	CDS	T→C	Phe261Leu
1894839	[a]pabB	CDS	T→C	Leu12Pro
2003346	[a]fliC	CDS	C→A	Asn87Lys
2040433	[a]yedY	CDS	C→A	Ala319Asp
2322251	atoC	CDS	T→G	Ile129Ser
3035546	prfB	CDS	A→G	Thr246Ala
3214757	[a]rpoD	CDS	T→C	Tyr571His
3300572	yhbS	CDS	G→A	Asp63Asn
3388041	[a]aaeB	CDS	A→C	Thr50Pro
3554135	[a]malT	CDS	T→A	Trp351Arg
3725176	[a]glyQ	CDS	A→C	Glu48Ala
4243857	[a]malF	CDS	G→T	Gly407Cys
4300405	mdtP	CDS	A→T	Gln209Leu
4342047	[a]melA	CDS	T→A	Leu46Gln
289241	yagI	Upstream	C→A (−79)	
289281	yagI	Upstream	TTGG→CTGA (−119)	
579146	nmpC	Upstream	T→C (−2321)	
579651	nmpC	Upstream	G→A (−2826)	
579671	nmpC	Upstream	A→G (−2846)	
579717	nmpC	Upstream	T→G (−2892)	
579811	nmpC	Upstream	G→A (−2986)	
687852	hscC	Upstream	C→A (−4459)	
696470	ybeX	Upstream	G→A (−4686)	
1299464	insZ	Upstream	A→C (−4142)	
1665170	clcB	Upstream	A→C (−145)	
1979271	cheA	Upstream	ATG→TTT (−3947)	
2118488	wcaN	Upstream	G→A (−4161)	
2118495	wcaN	Upstream	C→A (−4168)	
2118501	wcaN	Upstream	TGTGCTCGGGTCTT→AGGTCC (−4175)	
2118526	wcaN	Upstream	T→A (−4199)	
2118560	wcaN	Upstream	Insert T (−4233)	
2118599	wcaN	Upstream	TGTGCTCGGGACC→GCGTACAGATT (−4272)	
2118649	wcaN	Upstream	C→T (−4322)	
2725818	[a]kgtP	Upstream	T→C (−72)	
3707947	dppD	Upstream	G→T (−4099)	
4035734	fadB	Upstream	A→C (−4763)	
4166470	trmA	Upstream	G→A (−3200)	

(*continued on next page*)

Table 2 (*continued*)

Location	Gene	Type	CDNA change	AA change
4223638	arpA	Upstream	A→G (−1151)	
4296380	nrfD	Downstream	Insert CG (4948)	
4510238	yjhD	Upstream	A→C (−3382)	
4542681	nanM	Upstream	ΔA (−3917)	

Fs, frameshift.
[a] Mutation also reported for NCM3722 (Lyons et al., 2011).
[b] , stop codon.

During our evaluation of large genomic changes we found a 3.5 kb insert in LS5218 containing three genes putatively annotated for fatty acid and secondary metabolite biosynthesis. These genes have homology towards an acyl-carrier protein (B1R43_RS14595), an aldehyde/flavonoid reductase with an NAD(P) binding site (B1R43_RS14600) and a fatty acyl-AMP ligase (B1R43_14605). None of these enzymes have been studied, but their putative annotations suggest that they could augment fatty acid metabolism with new or enhanced enzymes. Among the small changes compared to MG1655, mutations in *rpoS* (sigma-28) and *rpoD* (sigma-70) could have large pleiotropic effects on the cell. LS5218 also has a mutation in *prfB* (release factor 2), similar to that of *E. coli* BL21. These mutations could explain the differences in gene expression (identified by microarray) between MG1655 and NCM3722, a close LS5218 relative (Soupene et al., 2003). This study showed significantly higher mRNA expression of flagella and chemotaxis and lower expression of galactitol and maltose operon and regulons (Soupene et al., 2003).

3. Conclusions

The genome sequence of *E. coli* LS5218 disproves a commonly held assumption about its relationship to the reference K12 strain. LS5218 is a close relative of NCM3722 and not MG1655, with many of the genomic differences reported here also seen in a comparison of the *E.*

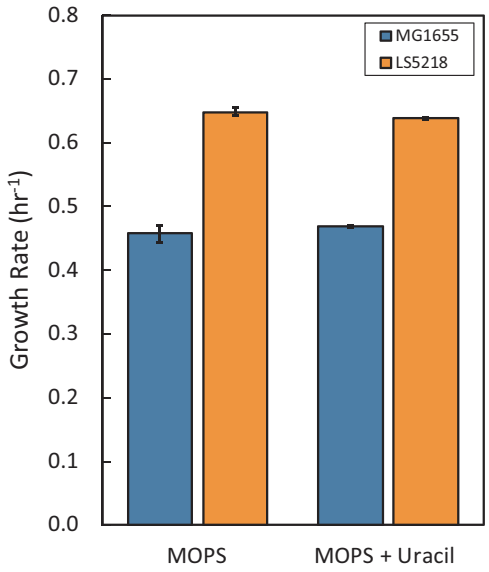

Fig. 3. *E. coli* MG1655 and LS5218 growth rate in MOPS minimal media with glucose or glucose supplemented with 20 μg/mL uracil.

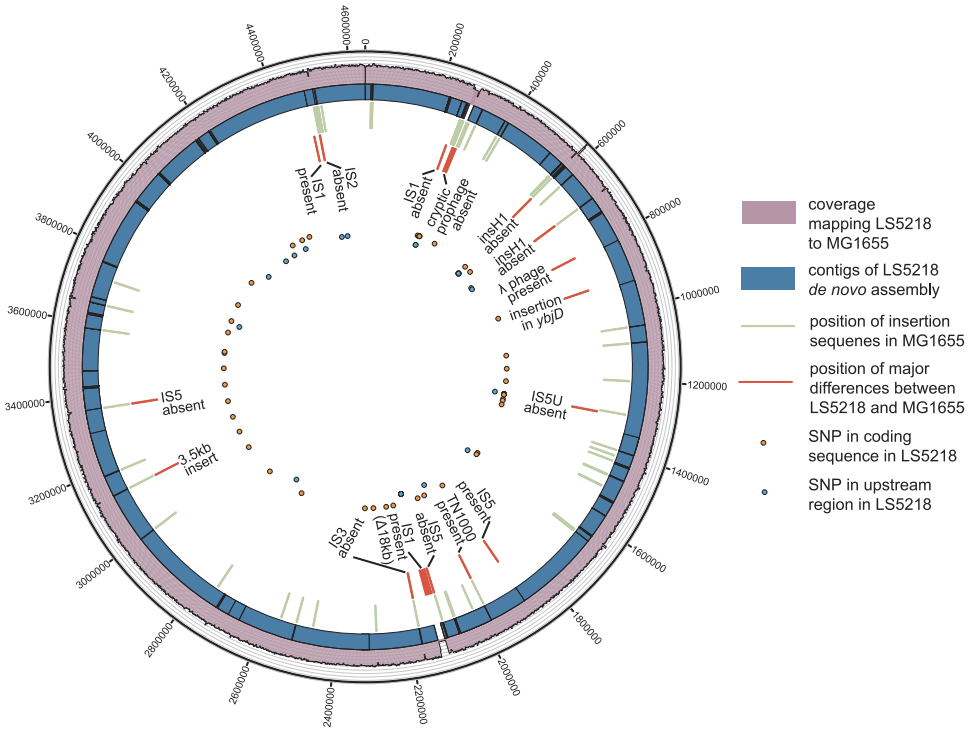

Fig. 2. Circular plot of LS5218 features mapped to MG1655. The outer purple histogram displays the coverage of LS5218 sequencing reads as mapped to MG1655. Average coverage was 200× with breaks displayed at genomic regions that differ between the strains. The contigs generated from the LS5218 *de novo* assembly are blue. Many of these breaks correspond to locations of native MG1655 insertion sequences (green bars). The large insertions and deletions of LS5218 are displayed in red and labeled. SNPs are spread throughout with those in coding regions shown in orange and those upstream of genes shown in light blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

coli strains MG1655 and NCM3722 (Tables 1, 2). Unfortunately the strain history for NCM3722 was lost (Lyons et al., 2011; Soupene et al., 2003) so we do not know if they are directly related. We theorize, based off phylogeny and common genetic variations, that NCM3722 and LS5218 share a similar derivation path and are a better representation of the original *E. coli* K-12 isolate than MG1655 (Soupene et al., 2003). The sequence of *E. coli* LS5218 allows us to have a better understanding of the genetic background for this widely used production strain and raises the question whether other mutations, in addition to *fadR601* and *atoC*(*c*), could be contributing to the improved production rates compared to other *E. coli* derivatives (Salamanca-cardona et al., 2014; Tappel et al., 2012; Ushimaru et al., 2015). The additional overlooked differences between LS5218 and MG1655 highlight the fact that genetic background is an important feature when selecting a host for metabolic engineering. The choice may have profound effects on successful engineering and strain performance.

4. Materials and methods

DNA was isolated from LS5218 using the Wizard® Genomic DNA Purification Kit (Promega) and sequenced by the University of Wisconsin Biotechnology Center. A paired end library was run on an Illuminia Hi-Seq. 2500. Sequencing generated 5431,968 reads (2 × 250). A *de novo* assembly was created using SPAdes (Bankevich et al., 2012). The draft genome contained 121 contigs (200 bp or greater) with an N50 of 159,470. The genome length was 4699,198 with an average coverage of 279X. The assembly included the complete F plasmid (67,502 bp) and bacteriophage phiX174 (5513 bp). The draft genome was annotated through the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). The genome sequence has been deposited in GenBank under bioproject PRJNA374891 and accession number MVJG00000000. Reads have been deposited to the Sequence Read Archive with accession number SRR5572609.

Sequencing reads (as FASTQ files) of *E. coli* LS5218 were mapped to completed reference genomes *E. coli K12 MG1655* (GCA_000005845.2) and *E. coli* NCM3722 (GCF_001043215.1) using Bowtie2 using the "fast-local" setting (Langmead and Salzberg, 2012). The output sequence alignment map (SAM) file was converted to a binary alignment map (BAM) file and sorted using SAMtools (Li et al., 2009). Variants were then called using FreeBayes (Garrison and Marth, 2012) and Naïve Variant Caller (Galaxy open source bioinformatics tool) (Goto et al., 2011). Variant calls were then annotated using SnpEff (Cingolani et al., 2012b) and variant calls with a quality of less than 30 were sorted out using SnpSift (Cingolani et al., 2012a). Large gaps and insertions were isolated using progressive Mauve alignment with default settings (Darling et al., 2004, 2010) and the pan-genome for the *E. coli* K-12 strains was generated with BPGA (default settings) (Chaudhari et al., 2016).

Specific growth rates calculated from growth curves generated in MOPS minimal media (Neidhardt et al., 1974) supplemented with 0.2 wt% glucose and 20 μg/mL uracil, when indicated. OD600 measurements were taken at 30 min intervals by a Tecan m200.

Acknowledgements

The authors thank the University of Wisconsin DNA Sequencing Facility for their sequencing services. G. Suen for assistance with bioinformatics analysis.

Funding statement

This work was funded by the National Science Foundation (CBET-114678) and the William F. Vilas Trust. JMR is the recipient of a NSF Graduate Research Fellowship (DGE-1256259). GCG is the recipient of a National Institutes of Health (NIH) Biotechnology Training Fellowship (NIGMS-5 T32 GM08349).

References

Agnew, D.E., Stevermer, A.K., Youngquist, J.T., Pfleger, B.F., 2012. Engineering *Escherichia coli* for production of C 12 – C 14 polyhydroxyalkanoate from glucose. Metab. Eng. 14, 705–713. http://dx.doi.org/10.1016/j.ymben.2012.08.003.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M. a., Pevzner, P. a., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997a. The Complete genome sequence of *Escherichia coli* K-12. Science 80–277, 1453–1462.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997b. The complete genome sequence of Escherichia coli K-12. Science 277, 1453–1462. http://dx.doi.org/10.1126/science.277.5331.1453.

Chaudhari, N.M., Gupta, V.K., Dutta, C., 2016. BPGA- an ultra-fast pan-genome analysis pipeline. Sci. Rep. 6, 24373. http://dx.doi.org/10.1038/srep24373.

Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., Lu, X., 2012a. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program. SnpSift. Front. Genet. 3. http://dx.doi.org/10.3389/fgene.2012.00035.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012b. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. (Austin). 6, 80–92. http://dx.doi.org/10.4161/fly.19695.

Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements, pp. 1394–1403. https://doi.org/10.1101/gr.2289704.

Darling, A.E., Mau, B., Perna, N.T., 2010. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5. http://dx.doi.org/10.1371/journal.pone.0011147.

Fujita, Y., Matsuoka, H., Hirooka, K., 2007. Regulation of fatty acid metabolism in bacteria. Mol. Microbiol. 66, 829–839. http://dx.doi.org/10.1111/j.1365-2958.2007.05947.x.

Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv Prepr. arXiv1207.3907 9. doi:arXiv:1207.3907 [q-bio.GN].

Goto, H., Dickins, B., Afgan, E., Paul, I.M., Taylor, J., Makova, K.D., Nekrutenko, A., Nekrutenko, A., Goto, H., Dickins, B., Afgan, E., Paul, I., Taylor, J., Makova, K., Nekrutenko, A., Chinnery, P., Thorburn, D., Samuels, D., White, S., Dahl, H., Turnbull, D., Lightowlers, R., Howell, N., Jacobs, H., DiMauro, S., Mercer, T., Neph, S., Dinger, M., Crawford, J., Smith, M., Shearwood, A.-M., Haugen, E., Bracken, C., Rackham, O., Stamatoyannopoulos, J., Filipovska, A., Mattick, J., Li, M., Wang, I., Li, Y., Bruzel, A., Richards, A., Toung, J., Cheung, V., Chen, R., Mias, G., Li-Pook-Than, J., Jiang, L., Lam, H., Chen, R., Miriami, E., Karczewski, K., Hariharan, M., Dewey, F., Cheng, Y., Clark, M., Im, H., Habegger, L., Balasubramanian, S., O'Huallachain, M., Dudley, J., Hillenmeyer, S., Haraksingh, R., Sharon, D., Euskirchen, G., Lacroute, P., Bettinger, K., Boyle, A., Kasowski, M., Grubert, F., Seki, S., Garcia, M., Whirl-Carrillo, M., Gallardo, M., Blankenberg, D., Taylor, J., Schenck, I., He, J., Zhang, Y., Ghent, M., Veeraraghavan, N., Albert, I., Miller, W., Makova, K., Hardison, R., Nekrutenko, A., Goecks, J., Nekrutenko, A., Taylor, J., Afgan, E., Baker, D., Coraor, N., Goto, H., Paul, I., Makova, K., Nekrutenko, A., Taylor, J., Marth, G., Korf, I., Yandell, M., Yeh, R., Gu, Z., Zakeri, H., Stitziel, N., Hillier, L., Kwok, P., Gish, W., Li, M., Schonberg, A., Schaefer, M., Schroeder, R., Nasidze, I., Stoneking, M., Bar-Yaacov, D., Avital, G., Levin, L., Richards, A., Hachen, N., Jaramillo, B.R., Nekrutenko, A., Zarivach, R., Mishmar, D., Nekrutenko, A., Taylor, J., Danecek, P., Auton, A., Abecasis, G., Albers, C., Banks, E., Depristo, M., Handsaker, R., Lunter, G., Marth, G., Sherry, S., McVean, G., Durbin, R., 2011. Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. Genome Biol. 12, R59. http://dx.doi.org/10.1186/gb-2011-12-6-r59.

Jenkins, L.S., Nunn, W.D., 1987. Genetic and molecular characterization of the genes involved in short-chain fatty acid degradation in *Escherichia coli*: the *ato* system. J. Bacteriol. 169, 42–52.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. http://dx.doi.org/10.1038/nmeth.1923.

Lawther, R.P., Calhoun, D.H., Gray, J., Adams, C.W., Hauser, C.A., Hatfield, G.W., 1982. DNA sequence fine-structure analysis of *ilvG* (IlvG+) mutations of *Escherichia coli* K-12. J. Bacteriol. 149, 294–298.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.

Lioliou, E.E., Mimitou, E.P., Grigoroudis, A.I., Panagiotidis, C.H., Panagiotidis, C.A., Kyriakidis, D.A., 2005. Phosphorylation activity of the response regulator of the two-component signal transduction system AtoS-AtoC in *E. coli*. Biochim. Biophys. Acta 1725, 257–268. http://dx.doi.org/10.1016/j.bbagen.2005.06.019.

Lyons, E., Freeling, M., Kustu, S., Inwood, W., 2011. Using Genomic Sequencing for Classical Genetics in *E. coli* K12. PLoS One 6. http://dx.doi.org/10.1371/journal.pone.0016717.

Matta, M.K., Lioliou, E.E., Panagiotidis, C.H., Kyriakidis, D.A., Panagiotidis, C.A., 2007. Interactions of the antizyme AtoC with regulatory elements of the *Escherichia coli atoDAEB* operon. J. Bacteriol. 189, 6324–6332. http://dx.doi.org/10.1128/JB.00214-07.

Nduko, J.M., Suzuki, W., Matsumoto, K., Kobayashi, H., Ooi, T., Fukuoka, A., Taguchi, S., 2012. Polyhydroxyalkanoates production from cellulose hydrolysate in *Escherichia coli* LS5218 with superior resistance to 5-hydroxymethylfurfural. JBIOSC 113, 70–72. http://dx.doi.org/10.1016/j.jbiosc.2011.08.021.

Neidhardt, F.C., Frederick, C., Curtiss, R., 1996. *Escherichia coli* and *Salmonella:* Cellular and Molecular Biology. ASM Press.

Neidhardt, F.C., Bloch, P.L., Smith, D.F., 1974. Culture medium for enterobacteria. J. Bacteriol. 119, 736–747.

Rhie, H.G., Dennis, D., 1995. Role of *fadR* and *atoC*(con) mutations in poly(3-hydroxybutyrate-co-3-hydroxyvalerate) synthesis in recombinant pha+ *Escherichia coli.* Appl. Environ. Microbiol. 61, 2487–2492.

Salamanca-cardona, L., Scheel, R.A., Lundgren, B.R., Arthur, J., Matsumoto, K., Taguchi, S., Nomura, C.T., Scheel, R.A., Lundgren, B.R., 2014. Deletion of the *pflA* gene in *Escherichia coli* LS5218 and its effects on the production of polyhydroxyalkanoates using beechwood xylan as a feedstock. Bioengineered 5, 284–287. http://dx.doi.org/10.4161/bioe.29595.

Simons, R.W., Egan, P.A., Chute, H.T., Nunn, W.D., 1980. Regulation of Fatty Acid Degradation in *Escherichia coli*: Isolation and Characterization of Strains Bearing Insertion and Temperature-Sensitive Mutations in Gene *fadR*, 142, pp. 621–632.

Soupene, E., Heeswijk, W.C., Van, Plumbridge, J., Stewart, V., Bertenthal, D., Lee, H., Prasad, G., Paliy, O., Charernnoppakul, P., Kustu, S., 2003. Physiological studies of *Escherichia coli* strain MG1655: growth defects and apparent cross-regulation of gene expression. J. Bacteriol. 185, 5611–5626. http://dx.doi.org/10.1128/JB.185.18.5611.

Spratt, S.K., Ginsburgh, C.L., Nunn, W.D., 1981. Isolation and genetic characterization of *Escherichia coli* mutants defective in propionate metabolism. J. Bacteriol. 146, 1166–1169.

Tappel, R.C., Wang, Q., Nomura, C.T., 2012. Precise control of repeating unit composition in biodegradable poly(3-hydroxyalkanoate) polymers synthesized by *Escherichia coli*. J. Biosci. Bioeng. 113, 480–486. http://dx.doi.org/10.1016/j.jbiosc.2011.12.004.

Theodorou, E.C., Theodorou, M.C., Samali, M.N., Kyriakidis, D.A., 2011. Activation of the AtoSC two-component system in the absence of the AtoC N-terminal receiver domain in *E. coli*. Amino Acids 40, 421–430. http://dx.doi.org/10.1007/s00726-010-0652-x.

Ushimaru, K., Watanabe, Y., Hiroe, A., Tsuge, T., 2015. Short Communication A Single-nucleotide Substitution in Phasin Gene Leads to Enhanced Accumulation of Polyhydroxyalkanoate (PHA) in Escherichia Coli Harboring Aeromonas Caviae PHA Biosynthetic Operon, 66, pp. 63–66. https://doi.org/10.2323/jgam.61.63.

van Aalten, D.M., DiRusso, C.C., Knudsen, J., Wierenga, R.K., 2000. Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold. EMBO J. 19, 5167–5177. http://dx.doi.org/10.1093/emboj/19.19.5167.

Xu, Y., Heath, R.J., Li, Z., Rock, C.O., White, S.W., 2001. The FadR·DNA complex. J. Biol. Chem. 276, 17373–17379. http://dx.doi.org/10.1074/jbc.M100195200.