Title.

An integrative approach points to membrane composition as a key factor in *E. coli* persistence

AUTHORS

Cañas-Duarte. SJ[1,2], Perez-Lopez. MI[1], Herrfurth. C[3], Sun. Lei[2], Contreras. LM[1], Feussner. I[3], Leidy. C[4], Riaño-Pachón, DM[5], Restrepo. S[6], and Pedraza. JM[4]

AFFILIATIONS

[1]Biological Sciences Department, Universidad de los Andes, Bogotá, Colombia, [2]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA, [3]Plant Biochemistry Department, Georg-August-University, Göttingen, Germany, [4]Physics Department, Universidad de los Andes, Bogotá, Colombia, [5]Computational, Evolutionary and Systems Biology Laboratory, Center for Nuclear Energy in Agriculture, Universidade de São Paulo, Piracicaba, Brazil, [6] Chemical and food Engineering Department, Universidad de los Andes, Bogotá, Colombia.

Summary:

In spite of its medical importance, the genetic mechanisms of bacterial persistence, particularly spontaneous (type II) persistence, remain largely unknown. We use an integrative approach, combining mutant genome analysis, transcriptomics and lipid membrane composition analysis, to elucidate said mechanisms. In particular, we analyzed the genome of the high persistence mutant *E. coli* DS1 (hipQ), to identify candidate mutations responsible for the high persistence phenotype. Contrary to a recent study, we find no mutation in ydcI. We compared the expression of spontaneous persistent and growing cells using RNAseq, and find that the activation of stress response mechanisms is likely less important in spontaneous persistence than recent reports suggest. It also indicated that modifications in the cell membrane could play an important role. This hypothesis was then validated by the analysis of the fatty acid composition of persister cells of both types, which have markedly different saturation from growing cells and between each other. Taken together, our results indicate that changing membrane composition might be a key process in persistence.

Highlights:

RNAseq analysis of spontaneous persistence shows no evidence of stress response

Identification of candidate SNPs for hipQ phenotype, excludes ydcI

Membrane fatty acid composition is involved in both types of bacterial persistence

Keywords: Persistence, spontaneous persisters, multi-drug tolerance, cell membrane modifications, stress response mechanisms.

Introduction.

Clonal populations of bacteria stochastically generate "persister" cells (Balaban et al., 2004; Hansen et al., 2012; Lewis, 2010), which exhibit transient multitolerance against bactericidal substances through a mechanism unrelated to acquired genetic resistance (Dörr et al., 2009). Persister cells are implicated in the development of chronic bacterial infections and may influence the appearance of resistance since persisters may be selected during antimicrobial therapy in response to infections (Hansen et al., 2012; Lewis, 2010). Persister cells have also been associated with the recalcitrance of biofilms (Lewis, 2001; Spoering and Lewis, 2001); making them a factor in the failure of antibiotic therapies to completely eradicate chronic bacterial infections (Hansen et al., 2012). Persistence is present among prokaryotic microorganisms such as *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and the etiologic agent of tuberculosis *Mycobacterium tuberculosis* (Dhar and McKinney, 2007; Keren et al., 2011; Lewis, 2010), making the study of bacterial persistence a public health priority.

Persister cells have arrested or reduced cell growth (Balaban et al., 2004) but survive treatment with bactericidal substances, potentially acting as a constitutive "insurance policy" for bacterial populations in case of catastrophic events (Balaban et al., 2004; Gefen and Balaban, 2009). In support of this hypothesis, Toxin-Antitoxin (TA) loci and stress response mechanisms are both reported to be involved in the generation of persister cells (Hong and Wang, 2012; Keren et al., 2004; Korch et al., 2003; Lewis, 2010; Maisonneuve et al., 2011; Shah et al., 2006). Although this phenomenon is fundamentally stochastic, it is now clear that the transition into and out of the persistent state also depends on the environment (Balaban et al., 2004; Gefen and Balaban, 2009). Furthermore, some of the important components of this dependency have been identified in *E. coli.* In particular, the conserved GRPase Obg, through the stringent alarmone (p)ppGpp, induces the expression of the toxin HokB, which in turn arrests growth by depolarizing the membrane (Verstraeten et al., 2015). It has been proposed (Gerdes, 2016) that this change in membrane potential induces a feedback through Rnase E. Another pathway for induction of persistence is through the TolC efflux pump (Pu et al., 2016), whose amount is distributed heterogeneously in the population (Pu et al., 2016) but also increased upon exposure to antibiotics (Viveiros et al., 2007).

These findings illustrate both how external signals can influence the transition to persistence and the variety of factors that determine the state. They also have in common an important connection with the membrane. Despite these advances, the specific mechanisms of multitolerance remain largely unknown (Hong and Wang, 2012; Kint et al., 2012). The variety of insults that they can survive

includes not only different types of antibiotics, but also alkaline and enzymatic lysis (Cañas-Duarte et al., 2014), and dormancy alone is not enough to exhibit the persistence phenotype (Orman and Brynildsen, 2013). This indicates that there is still much to be discovered about the persistence state itself and how it results in multitolerance.

To date, the existence of at least two types of persister cells has been proposed (Balaban et al., 2004). Each type exhibits similar multitolerance to antibiotics, but they differ in that triggered (type I) persisters are non-growing cells that appear to be generated upon passage to the stationary phase or other stresses, whereas spontaneous (type II) persisters are slow-growing cells that are generated continuously during exponential growth (Balaban et al., 2004). Studies on the high triggered persistence mutant *hipA7* (*E. coli* TH1269 strain) were crucial in understanding the stochastic nature of this phenomenon and in identifying several of the general characteristics of persistence known to date (Balaban et al., 2004; Korch et al., 2003; Moyed and Bertrand, 1983; Wolfson et al., 1990). These include the involvement of TA modules and stress response mechanisms (Keren et al., 2004; Kint et al., 2012; Korch et al., 2003; Lewis, 2010; Shah et al., 2006). However, the differences between triggered and spontaneous persister cells indicate that they might be generated by different mechanisms.

Previous persistence studies have mainly focused on triggered persister cells from the mutant strain *hipA7* (TH1269) (Balaban et al., 2004; Moyed and Bertrand, 1983). However, despite the existence of a strain with increased spontaneous persister cell frequencies, *E. coli* DS1 (*hipQ*) (Balaban et al., 2004; Wolfson et al., 1990), the genetic mechanisms involved have not been studied in depth and remain largely unknown. This is due to difficulties in differentially isolating spontaneous persisters (Cañas-Duarte et al., 2014) and the probable involvement of more than one mutation in the high persistence phenotype of this strain (Moyed and Bertrand, 1983). A recent study (Hingley-Wilson et al., 2020) presented interesting observations on phenotypic variability as it relates to persistence and attempted to determine the genes responsible for this phenotype. They proposed ydcI as the main candidate, but we find that they were mistaken as this gene is not mutated in DS1.

Two important caveats should be mentioned. It has been proposed that the cells that are persistent to a particular insult are not the same subset of the population that is persistent to a different insult (Allison et al., 2011). While plausible, the similarity in the number of survivors to many insults and the relation between the number of observed slow- or non-growing cells and the number of survivors make it unlikely, or at the very least would mean that the subsets are mostly overlapping. Nontheless,

if that hypothesis were true it would mean that most of this paper refers only to persisters to lythic insult. A second caveat is that we do not explicitly distinguish between persisters and viable but non culturable (VBNC) cells. These are cells that, as is the case for persisters, spontaneously or as a result of unfavorable conditions stop growth but remain viable and become multitolerant (Li et al., 2014) . The main difference is that these cells do not begin growth shortly after they are transferred to favourable media, but require either a long time or special stimuli to do so. It has been claimed that these cells are the same as persisters, but have been confused because of large numbers of non-viable cells (Kim et al., 2018), although this has been contested (Bamford et al., 2017). It has also been suggested that persisters and VBNC cells are simply different points on a continuum of "dormancy depth" (Ayrapetyan et al., 2015). We therefore believe that while the distinction could have implications for medical applications, it is not vital for the purpose of identifying the mechanisms of multitolerance.

In this study, we present an integrative analysis using genomic, transcriptomic and membrane composition analyses on both triggered and spontaneous persisters, which provides a basis for pinpointing the molecular mechanisms in further studies. The development of a protocol for persister cell isolation with the capability to distinguish between triggered and spontaneous persisters (Cañas-Duarte et al., 2014) allows this comparative analysis. This protocol allowed us to assess the transcriptomic profile of cells in the persistence state without the interference of the independent activation of stress response mechanisms by traditional, antibiotic-treatment based methods. Specifically, we tested whether the genetic mechanisms previously reported to be related to the induction of persistence, such as TA modules and stress response mechanisms, are also involved in the generation of spontaneous persister cells from the strain *E. coli* DS1 (*hipQ*) (Kint et al., 2012; Lewis, 2010). We also identified novel SNP mutations in the DS1 genome that we propose as candidates responsible for its high spontaneous persistence phenotype. Notably, a gene recently proposed to include the key mutation for enhanced persistence (McFadden), ydcI, was found not to be mutated in this strain. Additionally, the transcriptomic analysis of spontaneous persister cells showed that the activation of the SOS response is likely not involved in the spontaneous persistence physiological state, contrary to several reports (Dörr et al., 2009, 2010; Gurnev et al., 2012; Kint et al., 2012). Finally, our findings suggest that modifications in the physicochemical properties of the cell membrane could be related to the formation of persister cells and constitute an important mechanism for their multitolerance to bactericidal agents.

Materials and methods.

Bacterial strains and growth conditions.

The bacterial strains used in this work and their relevant characteristics are described in the Supplementary Information and are listed in table S1. Bacterial cultures were grown in Luria-Bertani (LB) medium at 37^0C and 200 rpm unless otherwise specified.

Escherichia coli DS1 genome sequencing, assembly and annotation.

Total DNA was purified from a stationary phase culture of *E. coli* DS1 using the GenElute Bacterial Genome Extraction Kit (Sigma-Aldrich, St. Louis, USA) according to the manufacturer's protocols. The DS1 genome was sequenced on an Illumina HiSeq 2000 instrument using the 2x90 paired-end technology with a 500 bp insert size at BGI, formerly the Beijing Genomics Institute, China.

FastQC (Babraham Bioinformatics, Cambridge, United Kingdom) was utilized to visually inspect quality metrics of the raw reads. Reads were clipped, quality trimmed and quality filtered (with a minimum read length of 60 bps and a quality threshold of 20) using Flexbar (Dodt et al., 2012).

Clean reads were then *de novo* assembled using the CLC assembly cell (CLC Bio, Aarhus, Denmark). PET scaffolding was performed using the SSAKE-based Scaffolding of Pre-Assembled Contigs after Extension (SSPACE) v2.0 (Boetzer and Pirovano, 2012). The PAGIT (Post Assembly Genome Improvement Toolkit) (Swain et al., 2012) was utilized for reference-guided contig extension using ABACAS (Assefa et al., 2009), PET gap closing was performed using IMAGE (Tsai et al., 2010) and the quality assessment of the assembly was made using iCORN (Otto et al., 2010). Finally, Gapfiller was used (Boetzer and Pirovano, 2012) to close the majority of the remaining gaps. DS1 genome annotation was performed using the Rapid Annotations using Subsystems Technology program (Aziz et al., 2008).

Whole genome SNP detection.

SNP calling was performed as described by the BROAD GATK Best practices guidelines (v3) using initially both *E. coli* MG1655 and DH10B genomes as references. Briefly, the sequenced small reads were mapped against each reference genome using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009), and the coverage depth was analyzed using the Genome Analysis Toolkit (GATK) (McKenna et al., 2010), obtaining a 94X coverage for each genome. Next, duplicates were marked using Picard [http://picard.sourceforge.net] before performing local re-alignments. For the base

quality recalibration step, we built a database of polymorphic sites in *E. coli* using the genomes of strains BL21, S88, 0127:H6 E2348/69, O42, ETEC H10407, DH10B and MG1655, employing progressive Mauve (Darling et al., 2010) for the multiple sequences alignment. Finally, SNPs in the genome of *E. coli* DS1 were called against each reference genome with GATK (DePristo et al., 2011).

Lastly, variants were called against the genome of the parental strain KLY (KL16 with a yfp-Cam cassette) (Fridman et al., 2014) using Geneious Prime. To analyze novel mutations common in all reference genomes, Mauve (Darling, 2004; Darling et al., 2010) was used to align the annotated reference genomes.

RNASeq analysis of spontaneous persister cells from exponentially growing E. coli K12 DS1.

To obtain enough RNA from the spontaneous persister cells for the RNASeq analysis, 6 replicate flasks, each containing 300 mL of LB media, were individually inoculated with 10 µL of an overnight culture of *E. coli* DS1 and incubated at 37^0C and 200 rpm until the culture reached an OD of 0.5. After reaching the desired OD, the bacterial cultures were pelleted, immediately frozen and stored at -30^0C. This was repeated until a total of 9.3 liters of exponentially grown cultures was processed for each of the two biological replicates.

To extract total RNA from spontaneous persister cells, persisters were first isolated by employing the protocol described in (Cañas-Duarte et al., 2014), and the RNA from the lysed non-persister cells was completely degraded using RNAse A (Sigma-Aldrich, St. Louis, USA) before proceeding with the total RNA extraction. The complete degradation of the RNA from the non-persister cells was determined with a gel electrophoresis of RNA extractions of the supernatant. Prior to the extraction of the total RNA from persisters, the pellets were washed three times.

Phenol-chloroform RNA extractions were performed in duplicate for both exponentially growing cells and DS1-(*hipQ*)-strain spontaneous persister cells from a pellet equivalent to an initial culture of 150 mL and 4.5 L for each biological replicate, respectively.

DNA degradation and total RNA purification were performed with a Qiagen RNeasy kit according to the manufacturer's protocols (Qiagen, Hilden, Germany). RNASeq on each sample was performed at BGI on an Illumina HiSeq 2000 instrument using 2x101 paired-end tags and strand-specific chemistry. Raw reads were processed as indicated above.

Transcriptome assembly, annotation, and analysis.

For spontaneous persisters and exponentially growing cells from the DS1 strain, we assembled the complete transcriptome using Trinity (Grabherr et al., 2011) for both *de novo* and genome-guided assemblies. Each transcriptome was then used for protein prediction and annotation of genes using Trinotate (Grabherr et al., 2011). The Trinotate pipeline includes a homology search to known sequence data using blastx and blastp (Altschul et al., 1990), the identification of protein domains using HMMER v3.0 (Finn et al., 2011), a prediction of signal peptides with singalP (Petersen et al., 2011) and tmHMM (Krogh et al., 2001), and several comparisons to curated annotation databases, such as EMBL, Uniprot, KEGG (Kanehisa et al., 2012), eggNOG (Powell et al., 2012), and Gene Ontology (Ashburner et al., 2000).

TopHat2 was employed to map all PET reads to the reference genome (Kim et al., 2013). Expression levels were presented as Fragments per Kilobase of exon per Million reads (FPKMs), and differential gene expression analyses were performed using both CuffDiff2 (Trapnell et al., 2010, 2013) and NOISeq (Tarazona et al., 2011) using the cutoffs of a p-value of 0.05 and a q of 0.9, respectively.

Analysis of overrepresented GO terms in differentially expressed genes during spontaneous persistence.

To analyze the biological functions that are significantly regulated during spontaneous persistence, we performed a Blast2GO (Conesa et al., 2005) analysis over the complete *E. coli* deduced proteome. We then tested for significant overrepresentation of GO terms in the groups of differentially expressed, overexpressed and underexpressed genes derived from CuffDiff (Trapnell et al., 2010)and NOISeq (Tarazona et al., 2011) analysis performed between *E. coli* DS1 exponentially growing cells and spontaneous persisters. The overrepresentation analysis was performed using Blast2GO DAVID 6.7 (Huang et al., 2009b, 2009a) Fisher's exact test, with a false discovery rate (FRD) correction to account for multiple testing (Benjamini-Hochberg test) and Bonferroni's score both with a threshold of <0.005.

RT-qPCR validation of detected differentially expressed genes.

From the group of differentially expressed genes, the expression profiles of 14 genes were chosen to be validated with RT-qPCR. The selection of these genes accounted for their biological function and the existence of previous reports of their relevance to triggered persistence (Keren et al., 2004; Lewis, 2010; Shah et al., 2006). The housekeeping genes *dxs* and *opgH* were used to normalize the data. For this analysis, a sample of the identical RNA that was sequenced and the RNA from a biological replica

for each condition were converted to cDNA prior to the qPCR. The cDNA synthesis and qPCR were performed with the DyNAmo SYBR Green 2-Step qRT-PCR Kit (Thermo Scientific, Waltham, USA).

The validation of the differentially expressed genes through qPCR was performed using the relative quantification method with a standard curve on a 7500 Fast qPCR Instrument (Applied Biosystems, Life Technologies, California, USA). The statistical analysis of the obtained data was done with REST (Pfaffl et al., 2002) using the Pair Wise Fixed Reallocation Randomization Test method (Pfaffl et al., 2002).

Determination of total fatty acids

Lipids were extracted from *E. coli* strains MG1655 and DS1 during exponential growth and stationary phase. Triggered and spontaneous persister cells were isolated from a stationary phase culture of *E. coli* TH1269 and an exponential phase culture of *E. coli* DS1, respectively, using a published protocol (Cañas-Duarte et al., 2014).

For each of the above-mentioned samples, lipids were extracted by pelleting 60 mL of each culture (or its equivalent in cell population for the triggered and spontaneous persister cells after the isolation protocol); the pellets were then frozen at -20^0 C overnight. Frozen pellets were then lyophilized overnight to remove any residual water. Afterwards, samples were dispersed in a chloroform/methanol/water (3:1:1 v/v) mixture and then vortexed every 15 minutes for 4 hours. After two days three separate phases are visually identified, where the top phase is an aqueous phase, the middle phase is a protein-rich phase, and the lower phase is the organic phase enriched in total lipids. The lower organic phase was drawn off by aspiration and collected into clean glass tube. All glass is cleaned using a piranha (sulfuric acid and hydrogen peroxide) protocol to remove all organic residues. Chloroform was then evaporated with a steady stream of gaseous N_2 to form a thin film at the bottom of the test tube, and the lipids were then stored at -20C. Extracted lipids were analyzed by gas chromatography.

Preparation of methyl esters of fatty acids (FAMEs) for analysis by gas chromatography/flame ionization detection (GC/FID) was performed as already described (Miquel and Browse, 1992). For acidic hydrolysis, 1 ml methanol/toluene (2:1, v/v) containing 2.75 % (v/v) H_2SO_4 (95-97 %) and 2 % (v/v) dimethoxypropan was added to the dry sample material. For later quantification of the fatty acids, 20 µg triheptadecanoate (Tri-17:0) were added and the sample was incubated for 1 h at 80 °C. To extract the resulting FAMEs, 200 µl of saturated aqueous NaCl solution and 2 ml of hexane were added. The hexane phase was dried under streaming nitrogen and re-dissolved with equal

volumes of water and hexane. The hexane phase was filtrated with cotton wool soaked with $NaSO_4$ and dried under streaming nitrogen. Finally the sample was re-dissolved in 10 µl acetonitrile for GC analysis performed with a Agilent (Waldbronn, Germany) 6890 gas chromatograph fitted with a capillary DB-23 column (30 m x 0.25 mm; 0.25 µm coating thickness; J&W Scientific, Agilent). Helium was used as carrier gas at a flow rate of 1 ml/min. The temperature gradient was 150°C for 1 min, 150 – 200°C at 8 K/min, 200–250°C at 25 K/min and 250°C for 6 min. For verification of the peak identities, an aliquot of the sample was analyzed by GC/ mass spectrometric detection (GC/MS) using a Agilent 5973 Network mass selective detector connected to the gas chromatograph as described above. The injection temperature was 220°C. The temperature gradient as well as the carrier gas was carried out as described for the GC analysis. Electron energy of 70 eV, an ion source temperature of 230 °C, and a temperature of 350°C for the transfer line were used. Mass detection was performed in scan mode in an m/z range of 50 to 650. Lipid extraction and analysis were performed for each of the above described samples with three biological replicates.

Results.

Novel SNPs identified in *E. coli* DS1 genome.

To identify the genetic mechanisms related to the high spontaneous persistence phenotype of *E. coli* DS1, we first sequenced the DS1 genome, obtaining 2.5 million clean reads with a minimum length of 60 bp. The average coverage in the *de novo* genome assembly was 82,85X, with a maximum coverage of 358X, estimated using Tablet (Milne et al., 2013). Following the de novo assembly, we obtained a single pseudo-molecule of 4567805 bp with 47 gaps and 531 N's (Table 1). After the annotation step, 4,554 genes were predicted.

The comparison of KLY, MG1655 and DH10B to reference genomes revealed a total of 153, 255 and 349 SNPs, respectively (Table 2). As none of the reference strains show a high persistence phenotype, and the DH10B strain presents persistence frequencies below the wild type levels (data not shown), we only considered the 112 novel SNPs that were common to the three reference genomes for further analyses (Supplementary material S2). The genomic analysis performed also showed the presence of the F plasmid integrated into the chromosome (Wolfson et al., 1990).

Several of the identified polymorphisms appear on genes whose functions were previously reported to be related with persistence, such as stress response mechanisms, DNA replication and catabolic processes of amino acids and carbon sources (Keren et al., 2004; Korch et al., 2003). We found a novel mutation in the *hipA* locus different from the previously characterized hipA7 mutation known

to confer a high persistence phenotype (Korch et al., 2003). This novel mutation does not appear to generate any significant conformational change in the protein (data not shown), although it is located near the active site of this kinase. We also identified various mutations in genes involved in the metabolism of lipids, cell wall homeostasis and transmembrane transport. As discussed in the next section, these results suggest the existence of modifications in cell envelope physicochemical properties in persister cells.

It is worth noting that gene ydcI, identified in by Hingley-Wilson and collaborators (Hingley-Wilson et al., 2020) as the gene responsible for the high spontaneous phenotype persistence, was found not to be mutated in strain DS1 by our genomic analysis. To corroborate this, we designed specific primers for the genes ydcI and ybaI and performed sanger sequencing. No mutations were detected in these genes. Furthermore, we tested the persister frequencies to ampicillin of the ΔydcI strain and found it to be close to wild type levels, even slightly lower (Supplementary material S3). In order to do this, we transferred the deletion from the KEIO collection (Baba et al., 2006) into our MG1655 background using P1 transduction.

Differentially expressed genes in spontaneous persisters were distinct from previously reported transcriptomes of persister cells.

Utilizing the paired end and strand-specific reads obtained during the RNASeq analysis, we assembled a complete gene catalog of expressed genes of *E. coli* DS1 cells during spontaneous persistence and exponential growth using 27,525,259 and 40,130,287 paired end reads, respectively, with an average length of 78 bps. For spontaneous persister cells, we assembled and annotated 5,768 transcripts; 5,748 of these transcripts displayed BLAST hits, whereas the number of transcripts was 4,393 for the exponentially growing cells.

A total of 301 differentially expressed genes were detected. Of these, 217 were found to be down-regulated and the remaining 84 up-regulated. In the list of up-regulated genes, we noticed the overall absence of genes related to stress response mechanisms that were previously reported to be involved in the generation of persister cells (Hong and Wang, 2012; Keren et al., 2004; Shah et al., 2006). Notably, genes participating in the SOS mechanism did not display differential regulation during spontaneous persistence. Several genes related to other stress-response mechanisms were found in the list of down-regulated genes (Supplementary material S4).

Several genes involved in lipid biosynthesis and lipid modification were found to be up-regulated during spontaneous persistence. This up-regulation is consistent with the identification of several

novel mutations in genes related to lipid pathways in the DS1 genome. As expected, several genes from TA modules were found to be differentially expressed during this type of persistence; surprisingly, their behavior varied from one TA module to another. We found that several antitoxins were down-regulated, such as, *hipB, dinJ* and *relB,* whereas others, such as pspB, were up-regulated. In particular, we found that the *tisB* toxin was down-regulated.

Lipid metabolism is associated with spontaneous persistence.

We analyzed the complete set of differentially expressed genes to detect whether some biological functions were significantly over- or underrepresented in this cluster. We found that genes related to regulation of translation, regulation of protein metabolic processes, several macromolecules biosynthetic processes, response to stimuli and response to stress were significantly overrepresented in the differentially expressed gene cluster, highlighting the importance of the regulation of these functions in spontaneous persister cells.

We separately analyzed the clusters of up-regulated and down-regulated genes to identify the biological functions and processes related to spontaneous persistence. In the down-regulated gene cluster during spontaneous persistence, we found that genes related to cell division, responses to stimulus and stressful conditions, synthesis of macromolecules, translation and overall homeostasis related processes were significantly overrepresented (Figure 1A).

When analyzing the cluster of up-regulated genes, we found that functions related to the intake of carbon sources and to general catabolic processes were overrepresented. Biological functions related to lipid metabolic processes were significantly overrepresented, suggesting that modifications in the cell membrane composition and physicochemical properties are important during spontaneous persistence (Figure 1B).

RT-qPCR validates all tested differentially expressed genes predicted by our RNASeq analysis.

We selected a subset of 12 genes (Supplementary material S5) out of 301 total genes found to be differentially expressed with the RNASeq analysis for validation: 4 and 8 from the up- and down-regulated cluster, respectively. Besides the differential expression, the selection criteria for this subset of genes included biological functions and their relation with mechanisms previously related to the persistence phenomenon (Hong and Wang, 2012; Keren et al., 2004; Shah et al., 2006).

To select appropriate reference genes for the RT-qPCR experiments, the expression profile of previously reported housekeeping genes in the RNASeq dataset was analyzed. The expression of the

genes *opgH* and *dxs* appeared to be suitable as RT-qPCR reference genes, whereas other commonly used housekeeping genes in *E. coli*, such as *recA* and *mdh*, showed variations in their abundances when comparing the *E. coli* DS1 strain during its spontaneous physiological state to its exponential growth.

All tested genes displayed similar expression profiles in the qPCR analysis, as reported by our RNASeq analysis (Supplementary material S6), including the housekeeping genes *opgH* and *dxs* (data not shown).

Identification of candidate substitutions related to differential gene expression and potentially responsible for the high spontaneous persistence phenotype of *E. coli* DS1.

Firstly, we identified 19 SNPs directly related to the differential expression of genes in spontaneous persistence (Supplementary Table S7). A non-synonymous mutation in *narZ*, which encodes for the nitrate reductase Z subunit α, is related with the up-regulation of the NarZWY operon. Significantly, we identified a non-synonymous mutation in the transcriptional terminator Rho which correlates with the differential expression of at least 16 genes from the putative Rho-dependent termination regulon (Peters et al., 2009). The other 17 mutations identified exhibit correlations between regulators like SoxR and the mutated genes.

We then analyzed each of the identified mutations in terms of biological function, and we assessed their possible relevance to persistence by investigating the transcriptomics data on persister cells combined with several known factors involved during their generation such as stress response mechanisms, DNA replication and cell division and catabolic processes of amino acids and carbon sources (Keren et al., 2004; Korch et al., 2003; Shah et al., 2006). In this analysis, we selected 11 non-synonymous mutations as candidates responsible for the high persistence phenotype of this strain (Table 3), whose function is discussed below.

Analysis of fatty acid profiles validates the occurrence of significant membrane modifications in persister cells.

The results of our RNA-Seq analysis of triggered and spontaneous persisters show that genes involved in fatty acid biosynthesis, normally involved in the modulation of fatty acid chemical structure, become differentially expressed, indicating a possible role of lipid composition of the membrane on persistence. Furthermore, previous observations by our group showed that each type of persister tolerates different concentrations of a lytic cocktail designed to induce high osmotic

stress and cell wall degradation (Cañas-Duarte et al., 2014). This led us to study the lipid composition of the different types of persisters.

Overall, the fatty acid composition in normally growing cells from different *E. coli* strains was highly conserved (Supplementary Table S8). We only observe minimal variation of fatty acid profiles among among the different physiological states for most cases (Figure 2). However, significant differences were found when comparing both triggered and spontaneous persister cells with all other samples from non-persister cells during both the exponential growth and stationary phases. In triggered persisters, the proportion of saturated lipids changed significantly, from less than the 50% of all membrane lipids in non-persisters form the stationary phase, to more than 68% in triggered persister cells. This change is expected to significantly increases the membrane rigidity of triggered persisters, which might be related with its non-growing phenotype. Interestingly, spontaneous persisters increase the proportion of unsaturated fatty acids in their membrane, from an average of 41% in non persisters to more than 58% in spontaneous persister cells, being consistent with an overall increase in membrane fluidity of the cell membrane for spontaneous persisters. No changes in the average fatty acid length was observed for neither triggered nor spontaneous persister cells compared to normal cells (Figure 3)

Discussion.

Despite numerous investigations covering persistence, little is known about the biochemical and molecular mechanism conferring multitolerance, the intermediate steps between the toxin-antitoxin system stochastic initiation and the establishment of the phenotype, and the role of external stresses in modulating this process. The recent development of an appropriate isolation method of naïve persister cells from bacterial populations that avoids the induction of stress response mechanisms and separates triggered and spontaneous persister cells (Cañas-Duarte et al., 2014) allowed us to obtain several insights into the genetic mechanisms behind the generation of persister cells: i) the identification of novel polymorphisms associated with spontaneous persistence; ii) the first expression profile of spontaneous persister cells; and iii) the first report of the involvement of lipid modifications in the phenomenon of persistence.

Since the identification and later characterization of the mutations in the HipA toxin gene, responsible for the hip phenotype of one of the only two high persistence strains (*hipA7* and *hipQ*) currently known (Balaban et al., 2004; Korch et al., 2003; Moyed and Bertrand, 1983; Wolfson et al., 1990), various studies have investigated "persistence genes" (Dörr et al., 2009, 2010; Hong and

Wang, 2012; Keren et al., 2004; Maisonneuve et al., 2011; Shah et al., 2006). Recent studies have demonstrated the importance of Toxin-Antitoxin modules in the generation of persisters, notably the activation of toxins HipA, TisB and MqsR (Dörr et al., 2010; Hong and Wang, 2012; Korch et al., 2003). The activation of several stress-response mechanisms has also been reported to induce persistence (Dörr et al., 2010; Hong and Wang, 2012; Lewis, 2010; Shah et al., 2006). However, we have recently found evidence that suggests that several stress response mechanisms only induce persistence if antibiotics were used during the isolation of persister cells (manuscript in preparation). Remarkably, until a recent paper (Hingley-Wilson et al., 2020) the identification of polymorphisms influencing the *hip* phenotype of *E. coli* DS1 (*hipQ* strain) had only been attempted once in 1983 using conjugative mapping (Moyed and Bertrand, 1983). Establishing the differences between triggered and spontaneous persister cells (Balaban et al., 2004; Cañas-Duarte et al., 2014) and using DS1, the only known strain enriched in spontaneous persister cells, the identification of the mutations responsible for its phenotype could be identified using next generation sequencing and transcriptomic analysis. This is particularly notable given that the difference between the two types was a functional definition.

The genomic analysis of the complete *E. coli* DS1 genome allowed the identification of several genes carrying mutations that could be related to the high frequencies of spontaneous persister cells generated by this strain. The analysis of all the mutations present in the genome of *E. coli* DS1 identified a cluster of 28 genes potentially involved in the generation of persister cells. Some of them have been previously characterized and some are reported in the current study. One of the most notable mutations found in the DS1 genome was a novel polymorphism in the coding sequence of the HipA toxin gene. A single point mutation in nucleotide 724 changed the cytosine (C) for a thiamine (T) producing a non-synonymous change in the protein, changing original Alanine residue to a Valine. This mutation was distinct from the hipA7 mutation that has been studied for several years after its identification as the mutation responsible for the hip phenotype of hipA7 strain (Korch and Hill, 2006; Korch et al., 2003; Moyed and Bertrand, 1983). Several additional mutations occurred in genes previously related to the induction of persistence in *E. coli*, such as for genes related to stress response mechanisms, biofilm formation, quorum sensing and catabolic processes (Dörr et al., 2010; Hansen et al., 2008; Keren et al., 2004, 2011; Korch et al., 2003; Lewis, 2008; Shah et al., 2006; Spoering and Lewis, 2001). Notably, the analysis of the DS1 genome supported the hypothesis that modifications to the lipids composing the cell membrane might influence or cause multitolerance.

Little is known about spontaneous persister cells, besides the fact that they continuously generate from an exponentially growing culture and are slow growers, different from the non-growing phenotype of triggered persisters (Balaban et al., 2004). To expand the knowledge on spontaneous persisters, we sequenced the complete transcriptome of spontaneous persisters using exponentially growing cells as the control. A total of 180 genes were differentially expressed during spontaneous persistence: 64% were down-regulated, and 36% were up-regulated. Several biological functions were found to be significantly regulated: genes encoding for the response to biotic stimulus and the metabolic processes related to carbohydrates were up-regulated, and genes encoding for cell division, protein synthesis and cell homeostasis processes were down-regulated. This regulation contradicts the idea that the mechanism of multitolerance is primarily because of metabolic inactivity (Kint et al., 2012; Lewis, 2010). Genes from Toxin-Antitoxin modules were also regulated during persistence, but their regulation appeared to be specific to each pair as we found both toxin and antitoxin genes among the up-regulated and down-regulated clusters. In particular, the TisB toxin was down-regulated during spontaneous persistence, in contrast to recent reports that showed TisB's ability to induce persistence (Dörr et al., 2010; Lewis, 2010)

Several genes related to stress response mechanisms were in the down-regulated cluster, whereas only one stress-related gene, *pspD*, was up-regulated. This regulation was contrary to that in triggered persistence and to previous studies (Keren et al., 2004; Kint et al., 2012; Lewis, 2010; Shah et al., 2006). Notably, all genes related to the SOS response were non-regulated, or were beyond our resolution power. This particular and striking difference between the current expression profile of spontaneous persister cells and all previous transcriptomic studies performed on persister cells could be caused by the inherent differences between spontaneous and triggered persister cells. However, the induction of the SOS response by antibiotics during the isolation of persister cells in all previous studies is likely to have caused significant alterations to the expression profiles of the isolated persisters. Our findings suggest that the induction of stress mechanisms, such as the stringent response, acidic shift and oxidative stress, only cause persistence induction if the SOS response was previously induced.

Expression information does not determine the genes that mediate the transition to the persistent state, but the data provide the basis for the discovery of the responsible genes. Indeed, correlating the information of the expression profile for spontaneous persisters and the discovered SNPs in the genome of *E. coli* DS1 with previously reported transcriptomic data, the number of candidate mutations for the hip phenotype of DS1 strain could be narrowed down to 11 candidate genes (Table

3). Amongst the candidate genes the general outer membrane porin *ompF*, which is known to mediate the entry of various antibiotics (Cohen et al., 1988; Ghai et al., 2018) was found to have suffered a deletion which resulted in a predicted frame shift, which might severely affect the functionality of this porin. A second interesting candidate mutation was found in the MltG protein (formerly yceG). MltG is an inner membrane enzyme with endolytic murein transglycosylase activity which has been proposed to terminate nascent peptidoglycan synthesis (Yunck et al., 2016) and its deletion has been reported to reduce sensitivity to ampicillin (Babu et al., 2011). Additionally, the non-synonymous mutations found in the genes *rho, rpoB, rseB, hflX* and *rsxG* could broadly affect gene expression via alteration of transcription factor activity, translation regulation and mRNA stability, which could also increase noise in gene expression favoring the entrance into alternate phenotypic states.

A notable result of our transcriptomics analysis was the overrepresentation of genes related to fatty acid metabolism and lipid modifications among the biological functions strongly up-regulated during spontaneous persistence. These results not only correlate with some of the mutations found in the DS1 genome, but also support previous findings about differences in the cell membrane of persister cells (Cañas-Duarte et al., 2014). The chemical structure of fatty acids determines the level of lipid packing, and influences the biophysical properties of bacterial membranes. For example, an increased presence of saturated fatty acids in bacteria tends to increase the level of lipid packing (Zhang and Rock, 2008), while the incorporation of cis-unsaturations and iso- and anti-iso methylations at the end of the lipid chain tend to reduce the level of lipid packing leading to an increase in membrane fluidity.

We analyzed the fatty acid composition of triggered and spontaneous persister cells and compared the composition with normally growing cells from the wild type strain *E. coli* MG1655 in stationary growth and from the *E. coli* DS1 strain in exponential growth, respectively. The fatty acid composition of persister cells differed significantly from that of the normally growing cells, specifically those in which phase they were isolated namely stationary growing cells for triggered persisters and exponentially growing cells for spontaneous peristers, displaying a likely increase in the overall membrane rigidity in triggered persistence whilst spontaneous persisters appear to be increasing their membrane fluidity. Given that the regulation of the physicochemical properties of the cell membrane is a known bacterial strategy to survive challenging environments (Zhang and Rock, 2008) and based on previous reports on the relationship between membrane fluidity and tolerance to certain antibiotic agents (Parsons and Rock, 2013), alterations in membrane composition and fluidity emerge as a strong candidate for the ultimate mechanism of multitolerance.

In summary, we exploited a newly developed isolation protocol (Cañas-Duarte et al., 2014)to characterize the genome of a spontaneous high persistence mutant and the transcriptomes of both persistence states. Our results question the importance of the stress response for the induction of persistence and clarify the metabolism of the persistence cells. We identify candidate mutations responsible for the high spontaneous persistence, and exclude a recently reported gene. Finally, we noted differences in the membrane lipid composition between persister and non-persister cells, opening a new avenue of research on the biochemical mechanism of multitolerance.

Acknowledgments.

We thank the laboratory of D. C. Hooper for retrieving the *E. coli* DS1 (hipQ). We also thank Dr. Thomas M. Hill for sending us strains TH1268 and TH1269 and Nathalie Balaban for sending us the strains MGY and DSY. We thank Dr. Johan Paulsson for critically reading the manuscript and acknowledge funding from Colciencias through grant 120465843511 and Convocatoria de Programas de Investigación 2012 of Universidad de los Andes.

References.

Allison, K.R., Brynildsen, M.P., and Collins, J.J. (2011). Metabolite-enabled eradication of bacterial persisters by aminoglycosides. Nature *473*, 216–220.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Assefa, S., Keane, T.M., Otto, T.D., Newbold, C., and Berriman, M. (2009). ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics *25*, 1968–1969.

Ayrapetyan, M., Williams, T.C., and Oliver, J.D. (2015). Bridging the gap between viable but non-culturable and antibiotic persistent bacteria. Trends Microbiol. *23*, 7–13.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., et al. (2008). The RAST Server: rapid annotations using subsystems technology. BMC Genomics *9*, 75.

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*.

Babu, M., Díaz-Mejía, J.J., Vlasblom, J., Gagarinova, A., Phanse, S., Graham, C., Yousif, F., Ding, H., Xiong, X., Nazarians-Armavil, A., et al. (2011). Genetic Interaction Maps in Escherichia coli Reveal Functional Crosstalk among Cell Envelope Biogenesis Pathways. PLoS Genet. *7*, e1002377.

Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). Bacterial persistence as a phenotypic switch. Science *305*, 1622–1625.

Bamford, R.A., Smith, A., Metz, J., Glover, G., Titball, R.W., and Pagliara, S. (2017). Investigating the physiology of viable but non-culturable bacteria by microfluidics and time-lapse microscopy. BMC Biol. *15*, 121.

Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. Genome Biol. *13*, R56.

Boetzer, M., Henkel, C. V, Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics *27*, 578–579.

Cañas-Duarte, S.J., Restrepo, S., and Pedraza, J.M. (2014). Novel protocol for persister cells isolation. PLoS One *9*, e88660.

Cohen, S.P., McMurry, L.M., and Levy, S.B. (1988). marA locus causes decreased expression of OmpF porin in multiple-antibiotic-resistant (Mar) mutants of Escherichia coli. J. Bacteriol. *170*, 5416–5422.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674–3676.

Darling, A.E., Mau, B., and Perna, N.T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One *5*, e11147.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping

using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

Dhar, N., and McKinney, J.D.J. (2007). Microbial phenotypic heterogeneity and antibiotic tolerance. Curr. Opin. Microbiol. *10*, 30–38.

Dodt, M., Roehr, J., Ahmed, R., and Dieterich, C. (2012). FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology (Basel). *1*, 895–905.

Dörr, T., Lewis, K., and Vulić, M. (2009). SOS response induces persistence to fluoroquinolones in Escherichia coli. PLoS Genet. *5*, e1000760.

Dörr, T., Vulić, M., and Lewis, K. (2010). Ciprofloxacin causes persister formation by inducing the TisB toxin in Escherichia coli. PLoS Biol. *8*, e1000317.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29-37.

Gefen, O., and Balaban, N.Q. (2009). The importance of being persistent: heterogeneity of bacterial populations under antibiotic stress. FEMS Microbiol. Rev. *33*, 704–717.

Gerdes, K. (2016). Hypothesis: type I toxin-antitoxin genes enter the persistence field-a feedback mechanism explaining membrane homoeostasis. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *371*, 20160189.

Gerdes, K., and Semsey, S. (2016). Microbiology: Pumping persisters. Nature *534*, 41–42.

Ghai, I., Bajaj, H., Arun Bafna, J., El Damrany Hussein, H.A., Winterhalter, M., and Wagner, R. (2018). Ampicillin permeation across OmpF, the major outer-membrane channel in Escherichia coli. J. Biol. Chem. *293*, 7030–7037.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Gurnev, P., Ortenberg, R., Dörr, T., Lewis, K., and Bezrukov, S. (2012). Persister-promoting bacterial toxin TisB produces anion-selective pores in planar lipid bilayers. FEBS Lett. *586*, 2529–2534.

Hansen, S., Lewis, K., and Vulić, M. (2008). Role of global regulators and nucleotide metabolism in antibiotic tolerance in Escherichia coli. Antimicrob. Agents … *52*, 2718–2726.

Hansen, S., Vulić, M., Min, J., and Yen, T. (2012). Regulation of the Escherichia coli HipBA toxin-antitoxin system by proteolysis. PLoS One *7*, e39185.

Hingley-Wilson, S.M., Ma, N., Hu, Y., Casey, R., Bramming, A., Curry, R.J., Tang, H.L., Wu, H., Butler, R.E., Jacobs, W.R., et al. (2020). Loss of phenotypic inheritance associated with ydcI mutation leads to increased frequency of small, slow persisters in Escherichia coli. Proc. Natl. Acad. Sci. *117*, 4152–4157.

Hong, S., and Wang, X. (2012). Bacterial persistence increases as environmental fitness decreases. Microb. … *5*, 509–522.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. *40*, D109-14.

Keren, I., Shah, D., Spoering, A., Kaldalu, N., and Lewis, K. (2004). Specialized Persister Cells and the

Mechanism of Multidrug Tolerance in Escherichia coli. J. … *186*, 8172–8180.

Keren, I., Minami, S., Rubin, E., and Lewis, K. (2011). Characterization and transcriptome analysis of Mycobacterium tuberculosis persisters. MBio *2*, e00100-11.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kim, J., Chowdhury, N., Yamasaki, R., and Wood, T.K. (2018). Viable but non-culturable and persistence describe the same bacterial stress state. Environ. Microbiol. *20*, 2038–2048.

Kint, C., Verstraeten, N., Fauvart, M., and Michiels, J. (2012). New-found fundamentals of bacterial persistence. Trends Microbiol. *20*, 577–585.

Korch, S.B., and Hill, T.M. (2006). Ectopic overexpression of wild-type and mutant hipA genes in Escherichia coli: effects on macromolecular synthesis and persister formation. J. Bacteriol. *188*, 3826–3836.

Korch, S.B., Henderson, T.A., and Hill, T.M. (2003). Characterization of the hipA7 allele of Escherichia coli and evidence that high persistence is governed by (p)ppGpp synthesis. Mol. Microbiol. *50*, 1199–1213.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

Lewis, K. (2001). Riddle of biofilm resistance. Antimicrob. Agents Chemother. *45*.

Lewis, K. (2008). Multidrug tolerance of biofilms and persister cells. Curr. Top. Microbiol. Immunol. *322*, 107–131.

Lewis, K. (2010). Persister cells. Annu. Rev. Microbiol. *64*, 357–372.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, L., Mendis, N., Trigui, H., Oliver, J.D., and Faucher, S.P. (2014). The importance of the viable but non-culturable state in human bacterial pathogens. Front. Microbiol. *5*, 1–1.

Maisonneuve, E., Shakespeare, L.J., Jorgensen, M.G., and Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. Proc. … *108*, 13206–13211.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

Milne, I., Stephen, G., Bayer, M., Cock, P.J.A., Pritchard, L., Cardle, L., Shaw, P.D., and Marshall, D. (2013). Using Tablet for visual exploration of second-generation sequencing data. Brief. Bioinform. *14*, 193–202.

Miquel, M., and Browse, J. (1992). Arabidopsis mutants deficient in polyunsaturated fatty acid synthesis. Biochemical and genetic characterization of a plant oleoyl-phosphatidylcholine desaturase. J. Biol. Chem. *267*, 1502–1509.

Moyed, H.S., and Bertrand, K.P. (1983). hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. J. Bacteriol. *155*, 768–775.

Orman, M.A., and Brynildsen, M.P. (2013). Dormancy is not necessary or sufficient for bacterial persistence. Antimicrob. Agents Chemother. *57*, 3230–3239.

Otto, T.D., Sanders, M., Berriman, M., and Newbold, C. (2010). Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. Bioinformatics *26*, 1704–1707.

Parsons, J.B., and Rock, C.O. (2013). Bacterial lipids: metabolism and membrane homeostasis. Prog. Lipid Res. *52*, 249–276.

Peters, J.M., Mooney, R.A., Kuan, P.F., Rowland, J.L., Keles, S., and Landick, R. (2009). Rho directs widespread termination of intragenic and stable RNA transcription. Proc. Natl. Acad. Sci. *106*, 15406–15411.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786.

Pfaffl, M.W., Horgan, G.W., and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. *30*, e36.

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T., et al. (2012). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res. *40*, D284-9.

Pu, Y., Zhao, Z., Li, Y., Zou, J., Ma, Q., Zhao, Y., Ke, Y., Zhu, Y., Chen, H., Baker, M.A.B., et al. (2016). Enhanced Efflux Activity Facilitates Drug Tolerance in Dormant Bacterial Cells. Mol. Cell *62*, 284–294.

Shah, D., Zhang, Z., Khodursky, A., Kaldalu, N., Kurg, K., and Lewis, K. (2006). Persisters: a distinct physiological state of E. coli. BMC Microbiol. *6*, 53.

Spoering, A.L., and Lewis, K. (2001). Biofilms and planktonic cells of Pseudomonas aeruginosa have similar resistance to killing by antimicrobials. J. Bacteriol. *183*, 6746–6751.

Swain, M.T., Tsai, I.J., Assefa, S.A., Newbold, C., Berriman, M., and Otto, T.D. (2012). A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat. Protoc. *7*, 1260–1284.

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. Genome Res. *21*, 2213–2223.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53.

Tsai, I.J., Otto, T.D., and Berriman, M. (2010). Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol. *11*, R41.

Verstraeten, N., Knapen, W.J., Kint, C.I., Liebens, V., Van den Bergh, B., Dewachter, L., Michiels, J.E., Fu, Q., David, C.C., Fierro, A.C., et al. (2015). Obg and Membrane Depolarization Are Part of a Microbial Bet-Hedging Strategy that Leads to Antibiotic Tolerance. Mol. Cell *59*, 9–21.

Viveiros, M., Dupont, M., Rodrigues, L., Couto, I., Davin-Regli, A., Martins, M., Pagès, J.-M., and Amaral, L. (2007). Antibiotic Stress, Genetic Response and Altered Permeability of E. coli. PLoS One *2*, e365.

Wolfson, J.S., Hooper, D.C., McHugh, G.L., Bozza, M. a, and Swartz, M.N. (1990). Mutants of Escherichia coli K-12 exhibiting reduced killing by both quinolone and beta-lactam antimicrobial agents. Antimicrob. Agents Chemother. *34*, 1938–1943.

Yunck, R., Cho, H., and Bernhardt, T.G. (2016). Identification of MltG as a potential terminase for peptidoglycan polymerization in bacteria. Mol. Microbiol. *99*, 700–718.

Zhang, Y.-M., and Rock, C.O. (2008). Membrane lipid homeostasis in bacteria. Nat. Rev. Microbiol. *6*, 222–233.

Tables and Figures

Total number of scaffolds	1
Sum (bp)	4567864
Total number of N's	881
Number of GAP's	49
Sum (bp) no N's	4566983

Table 1. Metrics of the *E. coli* DS1 genome assembly. Assembly of the DS1 genome was performed *de novo* using CLC, obtaining initially 104 contigs. The scaffolding process was performed using first SSPACE(Boetzer et al., 2011) with clean PE reads from the sequencing of both the genome and transcriptome to obtain 58 scaffolds. The genome of *E. coli* MG1655 was used to refine the assembly using the PAGIT pipeline(Swain et al., 2012). Finally, Gapfiller(Boetzer and Pirovano, 2012) was used to close gaps. The RAST(Aziz et al., 2008) annotation server was used to predict and annotate genes in the DS1 genome.

Type	Strain	Count	Percent/ (Total SNPs)
MISSSENSE	MG1655	132	43.40%
	NZ_CP008801	58	37.90%
	DH10B	122	73.05%
NONSENSE	MG1655	3	1.57%
	NZ_CP008801	1	0.65%
	DH10B	3	1.80%
SILENT	MG1655	109	52.60%
	NZ_CP008801	89	58.17%
	DH10B	224	64.20%

Table 2. Analysis of Single Nucleotide Polymorphisms (SNPs) detected in the genome of *E coli* K12 DS1. The genomes of the KL16 parental strain, the wild type strain MG1655 and the DH10B strain were used as a reference; we detected a total of 153, 255 and 349 SNPs, respectively**.**

Gene	Protein Effect	Function
orn	Substitution	Oligoribonuclease involved in mRNA decay pathway
katE	Substitution	Catalase HPII. RpoS regulated, induced in stationary phase and hyperosmotic stress
rseB	Substitution	anti-sigma factor stabilizing protein RseB. Negative regulator of SigmaE
rho	Substitution	Transcription termination factor Rho
pgi	Substitution	glucose-6-phosphate isomerase
hflX	Substitution	Heat shock-induced ribosome rescue factor
ompF	Frame Shift	Outer membrane porin
mltG (yceG)	Substitution	Endolytic murein transglycosylase
rsxG	Substitution	SoxR [2Fe-2S] reducing system protein
hipA	Substitution	Serine/threonine-protein kinase toxin
rpoB	Substitution	RNA polymerase subunit β

Table 3. Candidate mutations for the high spontaneous persister phenotype of the the DS1 strain.

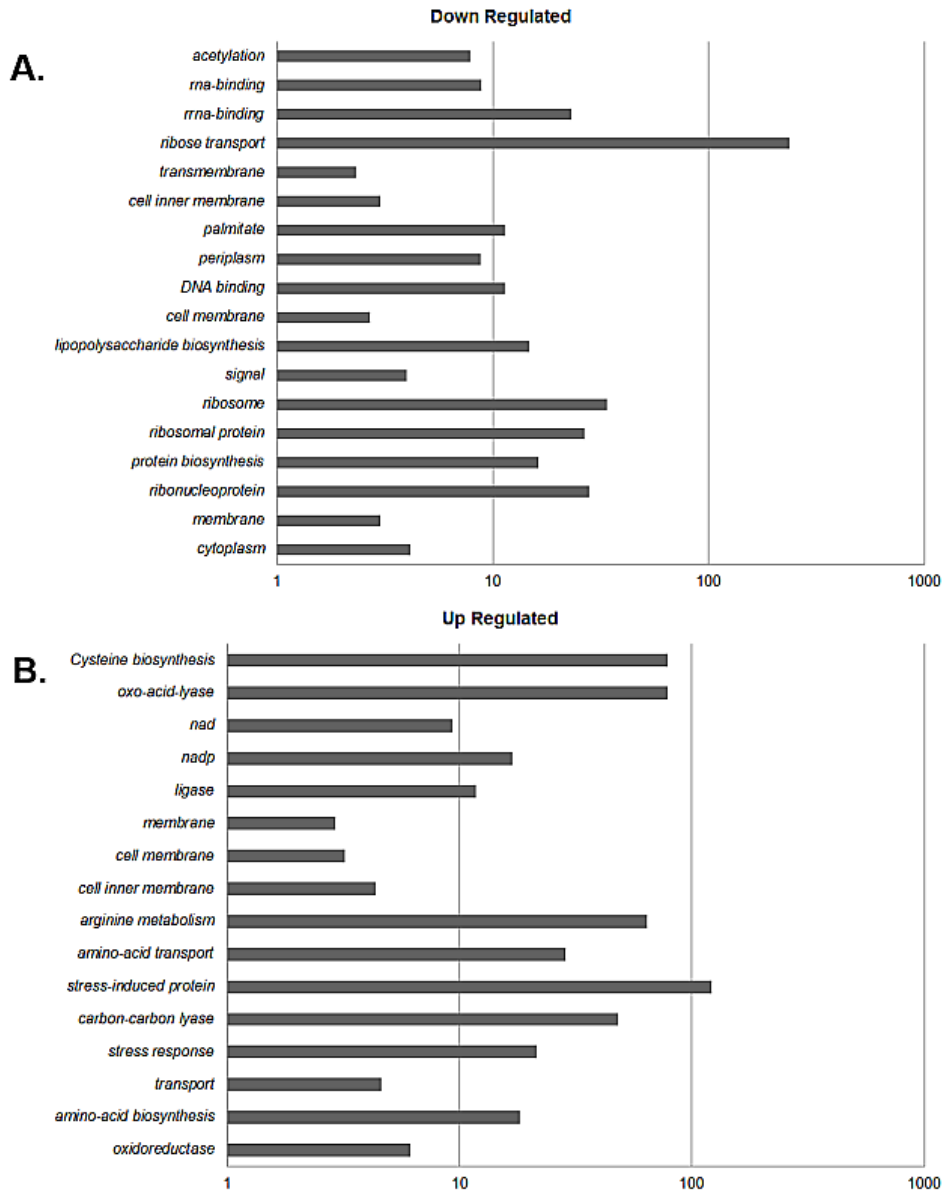

Figure 1. Biological Functions significantly overrepresented during spontaneous persistence. To analyze which biological functions being significantly regulated during spontaneous persistence, DAVID (Huang et al., 2009b, 2009a)was used to test for overrepresentation of biological functions in the subsets of A) down-regulated and B) up-regulated. As noted, several processes involving both DNA and RNA binding and metabolism of proteins are overrepresented in the set of down-regulated genes in spontaneous persister cells, whereas lipid metabolism is overall overrepresented un both the up-regulated and the down-regulated clusters, indicating a strong regulation occurring during spontaneous persistence in this function. Only biological functions found to be overrepresented with an Bonferroni score < 0.005 are shown and a FDR < 0.0005.

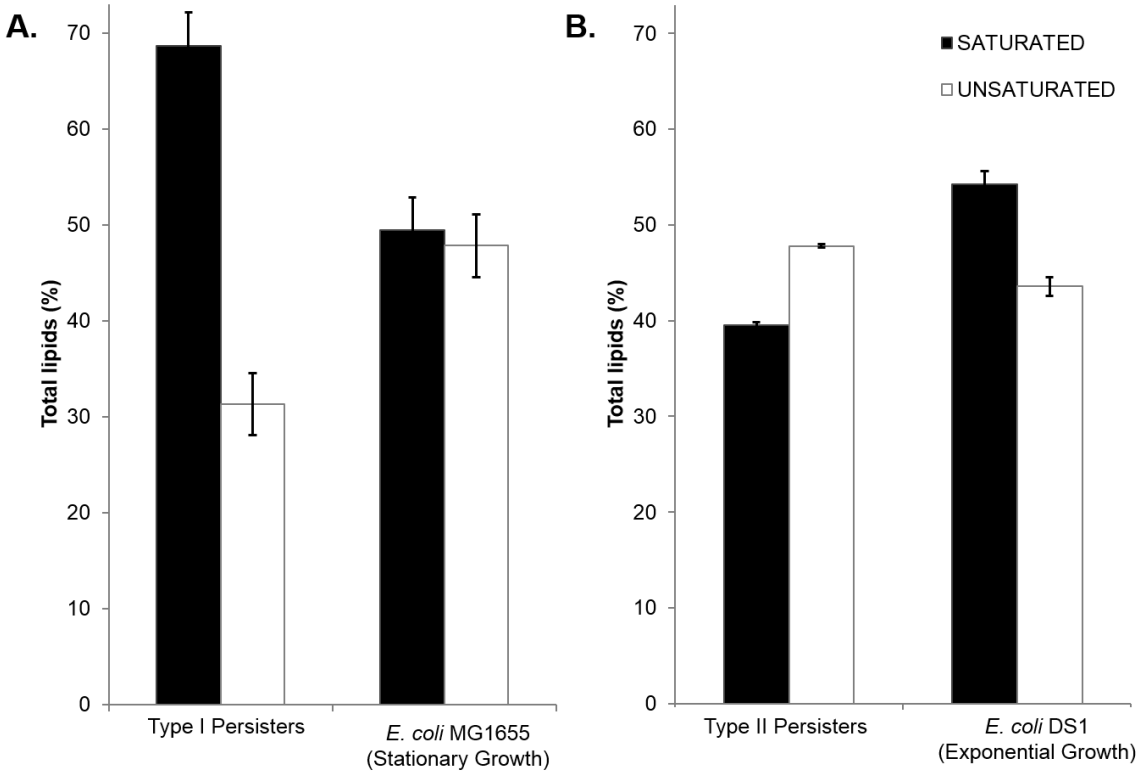

Figure 2. Fatty acid chain profiles of samples from different physiological states in *E. coli*; evidence of membrane modifications in triggered and spontaneous persister cells. The lipidic composition of persister cells is markedly different from that of the cells in the physiological state in which they are generated. **A)** Triggered persisters significantly increase their membrane rigidity by increasing the saturated lipids at the expense of reducing the amount of unsaturated lipids, whereas the membranes of **B)** spontaneous persisters are highly enriched in unsaturated lipids, consistent with an overall increase in the fluidity (decrease in lipid packing) of the cell membrane in persister cells.

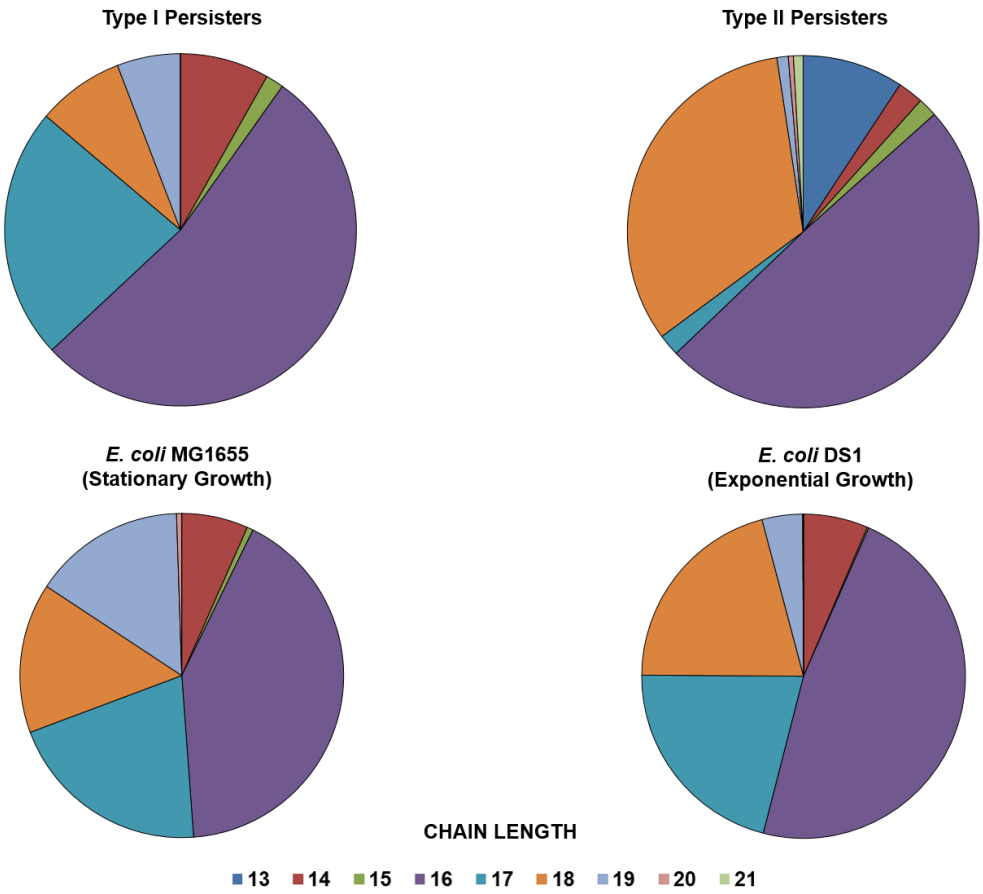

Figure 3. The fatty acid chain length profiles of persister cells is distinct from normally growing cells. The chain length of the lipids composing cell membranes of triggered and spontaneous persisters exhibit noticeable changes in composition. In particular, triggered persisters are enriched in both the 14 and 16 carbon chains, consistent with an increase in the saturated version of this species, compared with cells in stationary growth. Spontaneous persisters are enriched in 18 carbon chains, which is consistent with an increase in the unsaturaded 18:1Δ9 species. On average, the average chain length in both Triggered and Spontaneous is does not exhibit a significant change compared to normally growing cells, which is highly conserved amongst different physiological states and different *E. coli* strains.