Inter-replicon Gene Flow Contributes to Transcriptional Integration in the *Sinorhizobium meliloti* Multipartite Genome

George C. diCenzo, Deelaka Wellappili, G. Brian Golding, and Turlough M. Finan[1]
Department of Biology, McMaster University, Hamilton, Ontario, Canada L8S 4K1
ORCID IDs: 0000-0003-3889-6570 (G.C.d.); 0000-0002-7575-0282 (G.B.G.); 0000-0003-4363-6001 (T.M.F.)

ABSTRACT Integration of newly acquired genes into existing regulatory networks is necessary for successful horizontal gene transfer (HGT). Ten percent of bacterial species contain at least two DNA replicons over 300 kilobases in size, with the secondary replicons derived predominately through HGT. The *Sinorhizobium meliloti* genome is split between a 3.7 Mb chromosome, a 1.7 Mb chromid consisting largely of genes acquired through ancient HGT, and a 1.4 Mb megaplasmid consisting primarily of recently acquired genes. Here, RNA-sequencing is used to examine the transcriptional consequences of massive, synthetic genome reduction produced through the removal of the megaplasmid and/or the chromid. Removal of the pSymA megaplasmid influenced the transcription of only six genes. In contrast, removal of the chromid influenced expression of ∼8% of chromosomal genes and ∼4% of megaplasmid genes. This was mediated in part by the loss of the ETR DNA region whose presence on pSymB is due to a translocation from the chromosome. No obvious functional bias among the up-regulated genes was detected, although genes with putative homologs on the chromid were enriched. Down-regulated genes were enriched in motility and sensory transduction pathways. Four transcripts were examined further, and in each case the transcriptional change could be traced to loss of specific pSymB regions. In particularly, a chromosomal transporter was induced due to deletion of *bdhA* likely mediated through 3-hydroxybutyrate accumulation. These data provide new insights into the evolution of the multipartite bacterial genome, and more generally into the integration of horizontally acquired genes into the transcriptome.

Horizontal gene transfer (HGT) is pervasive in bacterial species and plays a significant role in the genotypic and phenotypic diversity between two isolates of the same species or genus (Ochman *et al.* 2000). The term 'pan-genome' is used to describe all genes present in all strains of a particular species or genus, with the 'core genome' referring to genes conserved in all strains and the 'accessory genome' used in reference to genes absent in at least one strain (Tettelin *et al.* 2005). The pan-genome of 33 *Sinorhizobium meliloti* strains was determined to consist of 21,000 genes of which 78% were variably present or absent (Sugawara *et al.* 2013). In the case of the genus *Prochlorococcus*, the genome size is a mere 2,000 genes in comparison to a pan-genome size of at least 58,000 genes (Baumdicker *et al.* 2012). Most genes acquired through HGT are rapidly lost from the genome (Lawrence and Ochman 1998; Hao and Golding 2006). Many factors influence whether the maintenance of a gene acquired through HGT is selected for, among which are that the gene must be appropriately expressed and regulated, and that it does not interfere with the expression or activity of existing gene products (Pál, Papp, and Lercher 2005b; Sorek *et al.* 2007; Amorós-Moya *et al.* 2010; Cohen *et al.* 2011; Park and Zhang 2012; Baltrus 2013; San Millan *et al.* 2015). *In silico* analyses have predicted that newly acquired genes are poorly integrated into existing regulatory networks, but that these genes are eventually integrated into the networks (Lercher and Pál 2008). However, to date there has been little experimental analysis of the effects of large-scale horizontal gene transfer on the existing transcriptome of the host cell.

Approximately 10% of bacterial species contain a multipartite genome consisting of at least two DNA replicons of 300 kb or larger (diCenzo and Finan 2017). In these genomes, there is a primary chromosome that contains most if not all core cellular processes and is primarily vertically transmitted (Harrison *et al.* 2010; Galardini *et al.* 2013). The secondary replicons can be broadly classified as megaplasmids or chromids. Megaplasmids originate through recent HGT and are continuously undergoing rapid gene gain and gene loss (Guo *et al.* 2009; Cooper *et al.* 2010; Galardini *et al.* 2013). As such, the characteristics of entire megaplasmids largely mimics those of individual genes acquired through recent HGT. Chromids can be thought of as domesticated megaplasmids that are largely derived from ancient HGT (Wong and Golding 2003; Harrison *et al.* 2010; Galardini *et al.* 2013). These replicons therefore display many of the same characteristics as individual genes obtained through ancient HGT events. In addition, chromids encode a few core processes as a result of inter-replicon gene flow from the chromosome (Harrison *et al.* 2010). A clear example of this is in the genus *Sinorhizobium*, where the translocation of a contiguous 69 kb region from the chromosome to the chromid resulted in the transfer of two genes essential for cell viability (diCenzo *et al.* 2013; diCenzo *et al.* 2016b). Studying the features of secondary replicons and how the cell responds to the gain or loss of these replicons can serve as a tool to study the effects of large-scale HGT and how the time since acquisition of horizontally transferred genes influences their properties.

In some ways, secondary replicons function as separate entities from the chromosome. Transcriptional studies with *S. meliloti*, *Burkholderia cenocepacia*, *Vibrio cholerae*, *Rhizobium leguminosarum*, and *Rhizobium phaseoli* indicated that genes from particular replicons were over-represented among the genes differentially expressed in response to adaptation to certain niches (Xu *et al.* 2003; Becker *et al.* 2004; Barnett *et al.* 2004; Yoder-Himes *et al.* 2010; Ramachandran *et al.* 2011; López-Guerrero *et al.* 2012). *In silico* metabolic modeling work with *S. meliloti* predicted that the chromid had a much greater contribution to fitness in the rhizosphere than in bulk soil (diCenzo *et al.* 2016a). An *in silico* analysis of regulons in *S. meliloti* predicted that most transcription factors regulate genes on their own replicon (Galardini *et al.* 2015).

On the other hand, inter-replicon co-operation does occur, and the level of integration appears dependent on how long the replicons have co-existed within the same cell. Despite occupying distinct territories of the nucleoid, there are physical interactions between the two replicons of *V. cholerae* (Galli *et al.* 2016) and initiation of chromid replication is dependent on a chromosomally encoded factor (Baek and Chattoraj 2014; Val *et al.* 2016). Elimination of the chromid from *S. meliloti* resulted in a major perturbation of the intra-cellular and extra-cellular metabolomics profile unlike elimination of the megaplasmid (Fei *et al.* 2016), although many changes may have been indirect effects of the impaired growth. An *in silico* analysis predicted functional links between each of the seven replicons of *Rhizobium etli*, with the two most recently acquired replicons having the least connections (González *et al.* 2006). Moreover, experimental and *in silico* studies suggest a minority of *S. meliloti* transcription factors regulate genes on other replicons, with greater cross-talk between the two older replicons (Pini *et al.* 2015; Galardini *et al.* 2015). Similarly, the chromosomally encoded RpoS regulates expression of both chromosomal and chromid genes in *V. cholerae* (Heidelberg *et al.* 2000). However, it has not been experimentally examined on a global scale the level of transcriptional integration of each replicon in a multipartite genome.

The genome of *S. meliloti* is split between a 3.65 Mb chromosome, a 1.68 Mb pSymB chromid, and a 1.35 Mb pSymA megaplasmid (Galibert *et al.* 2001), and derivatives that lack all of pSymA, pSymB, or both (a 45% genome reduction) have been constructed (Oresnik *et al.* 2000; diCenzo *et al.* 2014). Here, we use RNA sequencing to examine the transcriptional consequences of the loss of the chromid and/or the megaplasmid. Our data indicate that extended evolution of a chromosome and a secondary replicon in the same cell results in significant integration at the transcriptional level, and that this is partly driven by inter-replicon gene flow. These results provide novel insights into the evolution of multipartite genomes as well as into the integration of horizontally acquired genes into the genome.

MATERIALS AND METHODS

Growth conditions

Antibiotic concentrations, growth conditions for *S. meliloti* and *E. coli*, and the compositions of LB and LBmc were as described previously (diCenzo *et al.* 2014). LB and LBmc were supplemented with 2 μM $CoCl_2$ for growth of *S. meliloti* strains (Cheng *et al.* 2011). M9-sucrose minimal medium consisted of: 41 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 41 μM biotin, 42 nM $CoCl_2$, 1 mM $MgSO_4$, 0.25 mM $CaCl_2$, 38 μM $FeCl_3$, 5 μM thiamine-HCl, and 10 mM sucrose. When included, DL-3-hydroxybutyrate was added to a final concentration of 10 mM.

Genetic manipulations and bacterial strains

DNA manipulations and recombinant techniques, bacterial matings, and ΦM12 transductions were performed as described before (Finan *et al.* 1984; Sambrook *et al.* 1989; Cowie *et al.* 2006; Milunovic *et al.* 2014). Recombinant plasmids were produced through sequence and ligation independent cloning (Jeong *et al.* 2012). All bacterial strains and plasmids are listed in Table S1, and oligonucleotides are listed in Table S2.

All strains used for RNA-seq analysis were constructed previously and are listed in Table 1 (diCenzo *et al.* 2014; diCenzo *et al.* 2016b). Briefly, the two essential genes of pSymB (tRNA[arg] and *engA*) were integrated into the chromosome of a *S. meliloti* strain lacking pSymA, following which pSymB was removed to produce the strain ΔpSymAB (RmP3496) (diCenzo *et al.* 2014). The pSymA megaplasmid was conjugated into this strain to produce ΔpSymB (RmP3497) (diCenzo *et al.* 2016b). Similarly, ΔpSymA and the 'wild type' strains were produced by conjugating either pSymB or both pSymA and pSymB into the ΔpSymAB strain (diCenzo *et al.* 2016b). Finally, ΔpSymAB$_{\Omega NGR69}$ was constructed by integrating the 69 kb ETR region from *Sinorhizobium fredii* NGR234 into the chromosome of a *S. meliloti* strain lacking pSymA, followed by the removal of pSymB (diCenzo *et al.* 2016b). The ETR region (standing for *engA*-tRNA[arg]-*rmlC*) refers to a 69 kb region that translocated from the chromosome to pSymB in a *S. meliloti* ancestor (diCenzo *et al.* 2013).

RNA sequencing and analysis

Cell pellets from biological duplicates of each of the five strains (ΔpSymAB, ΔpSymB, ΔpSymA, wild type, ΔpSymAB$_{\Omega NGR69}$) grown in 5 mL of M9-sucrose were washed with fresh M9-sucrose and used to inoculate 100 mL M9-sucrose cultures to a starting OD$_{600}$ ∼0.1. Cultures were grown at 30° with shaking until an OD$_{600}$ ∼0.7 was reached, at which point 45 mL of cultures was immediately mixed with 5 mL of ice-cold cell stop solution (5% unbuffered phenol, 95% ethanol). Cells were pelleted by centrifugation at 3,700 *g* at 4° for 10 min, and the pellets flash frozen with liquid nitrogen. Frozen pellets were stored at -80° until use.

Isolation of total RNA from each pellet, as well as the confirmation of RNA integrity and lack of DNA contamination, was performed as

Strain	Identifier	Genotype	Reference
ΔpSymAB	RmP3496	A Rm2011 derivative with the two essential pSymB genes (engA and tRNAarg) in the chromosome, and lacking the pSymA megaplasmid and the pSymB chromid.	(diCenzo et al. 2014)
ΔpSymB	RmP3497	ΔpSymAB with the pSymA megaplasmid re-introduced into the genome	(diCenzo et al. 2016b)
ΔpSymA	RmP3498	ΔpSymAB with the pSymB chromid re-introduced into the genome	(diCenzo et al. 2016b)
Wild type	RmP3499	ΔpSymAB with the pSymA and pSymB replicons both re-introduced into the genome	(diCenzo et al. 2016b)
ΔpSymAB$_{\Omega NGR69}$	RmP3500	A Rm2011 derivative containing the entire ETR (engA-tRNA-rmlC) region of S. fredii NGR234 integrated into the chromosome, and lacking the pSymA megaplasmid and the pSymB chromid.	(diCenzo et al. 2016b)

described previously (diCenzo et al. 2017). Depletion of the rRNA with Ribo-Zero rRNA Removal kits (Illumina), cDNA library synthesis, and cDNA sequencing using an Illumina HiSequation 1500, with 60 bp single reads, was performed at the Farncombe Family Digestive Health Research Institute located at McMaster University.

As a quality control step, FastQC (Andrews 2010) was run on all sequencing samples. Reads were trimmed with cutadapt (Martin 2011) with options to ensure quality of 25 from both ends, minimum size of 20 nts, and no Ns at the ends. Each replicon of the S. meliloti Rm2011 genome (Sallet et al. 2013) were individually indexed with Burrows-Wheeler Aligner (bwa) (Li and Durbin 2009). The bwa algorithm was then used to index and map the reads to each replicon, and htseq-count (Anders et al. 2015) was used to count the number of reads to each annotated gene. Differential expression analysis was finally performed using DESeq2 (Love et al. 2014), with the DESeq2 analysis performed independently for each of the three replicons. The complete set of results from the DESeq2 analysis is summarized in File S2, File S3, and File S4, and the raw sequencing reads are available online through the Gene Expression Omnibus (GEO accession: GSE106129).

Computational analyses

Heatmaps were generated with the heatmap.2 function in the gplots package of R (Warnes et al. 2009), and the clustering performed using average linkage with a Pearson correlation distance. Operon prediction was performed with the Rockhopper software (Tjaden 2015) using the raw RNA-seq data generated in this study (File S5). Genes were classified as core, accessory, or dispensable based on a previously published pangenome analysis (Galardini et al. 2015). The Blast 2.6.0+ software (Camacho et al. 2009) was used to search the entire S. meliloti Rm2011 chromosomal proteome (NC_003047) against the pSymA proteome (NC_003037) and pSymB proteome (NC_003078) (Sallet et al. 2013). Putative homologs were identified as those that shared at least 30% identity over at least 60% of the query protein. All chromosomal genes were functionally annotated with Gene Ontology (GO) terms, Cluster of Orthologous Genes (COG) categories, and KEGG pathway terms using the eggNOG webserver (Ashburner et al. 2000; Tatusov et al. 2003; Huerta-Cepas et al. 2016; Kanehisa et al. 2016). Custom perl scripts were used to summarize the functional annotation data, and functionally similar GO terms were grouped using the REVIGO webserver (Supek et al. 2011).

Reporter enzyme assays

Biological triplicates of each strain of interest were grown in either 5 mL M9-sucrose test tube cultures or in 200 μL M9-sucrose cultures in 96-well microtitre plates. β-Galactosidase and β-glucuronidase assays were performed in 96-well microtitre plates as described elsewhere

(Cowie et al. 2006). The sole change to the procedure was that 10 μL of culture, 40 μL of reaction buffer, and 50 μL of stop buffer was used per reaction. Control reactions were prepared for each culture by mixing 40 μL start buffer and 50 μL stop buffer prior to the addition of 10 μL of culture. The A_{420} (β-galactosidase) or A_{405} (β-glucuronidase) of each reaction and control reaction, and the OD_{600} of each culture, was measured using a Biotek Cytation3. Miller Units were calculated following the subtraction of the A_{420}/A_{405} values of the control reaction from the A_{420}/A_{405} reading of the true reaction.

Confirmation of RNA-seq expression data

For confirmation of the RNA-seq data, lacZ or gusA transcriptional fusions to ten genes of interest were identified in the S. meliloti fusion library (Cowie et al. 2006). The fusion library consists of a collection of S. meliloti strains with pTH1522 plasmid single cross-over recombinants integrated at known locations throughout the S. meliloti genome (http://info.mcmaster. ca/fusionlibrary/sinoseq.php). The fusion strains used in this study were: SmFL55 (smc02350$^+$::lacZ), SmFL107 (smc01946$^+$::gusA), SmFL1109 (sma2085$^+$::lacZ, sma2087::gusA), SmFL1435 (smc03027$^+$:: lacZ), SmFL1485 (sma0089$^+$::gusA), SmFL1705 (smc03133$^-$::gusA), SmFL4407 (smc03121$^+$::lacZ), SmFL6033 (smc02780$^+$::gusA), and SmFL6082 (smb21039$^-$::lacZ). Each fusion was recombined via transduction from the appropriate fusion library stain into each of the five strains used in the RNA-seq analysis, selecting for the gain of gentamicin resistance.

Transcriptional changes in pSymB deletion mutant strains

To identify pSymB gene regions that modulate expression of four genes of interest (smc02350, smc03121, sma2085, and sma2087), the appropriate gene fusions were combined with the following eleven S. meliloti strains with large, defined deletions of pSymB (Milunovic et al. 2014): ΔB154 (pSymB nt. 62,137-100,636), ΔB141 (101,396-466,499), ΔB163 (451,557-651,863), ΔB180 (635,940-869,642), ΔB181 (870,505-1,129,758), ΔB108 (1,131,168-1,169,073), ΔB109 (1,180,566-1,204,770), ΔB179 (1,207,052-1,322,226), ΔB118 (1,323,078-1,528,150), ΔB124 (1,529,711-1,677,882), and ΔB161 (pSymB nt. 1,679,723-49,523) (Figure S1). In the case of the ΔB154, ΔB163, ΔB181, ΔB108, ΔB118, ΔB124, and ΔB161 deletions, the deletions were recombined via transduction of neomycin resistance into each of the fusion library strains of interest. In the case of the ΔB141, ΔB180, and ΔB109 deletions, the gene fusions of interest were recombined via transduction of gentamicin resistance into each deletion mutant strain. For the ΔB179 deletion, the gene fusions of interest were first recombined via transduction into S. meliloti RmP2719, which carries the essential engA and tRNA^{arg} genes integrated into the chromosome

(diCenzo *et al.* 2013), following which the ΔB179 deletion was recombined via transduction of neomycin resistance into the resulting strains.

The deletions ΔB110 (1,207,052-1,255,032), ΔB116 (1,256,503-1,307,752), and ΔB158 (1,308,912-1,322,226) were combined with the *smc03121::lacZ* gene fusion using the method described for ΔB179 above. Four new pSymB deletion mutants were constructed as follows. The plasmid pTH1943 (Milunovic *et al.* 2014), containing a *FRT* (Flippase Recognition Target) site, was conjugated into four fusion library strains (SmFL2626, SmFL2987, SmFL4163, SmFL4484) that each contained a *FRT* site (Cowie *et al.* 2006), and neomycin resistant transconjugants were purified. The plasmid pTH2505 (Zhang *et al.* 2012), which expresses the Flp recombinase, was conjugated to each of the four transconjugants resulting in the deletion of the region flanked by the *FRT* sites. This produced the deletions ΔB182 through ΔB185. Each deletion was then combined with the *smc03121::lacZ* gene fusion using the method described for ΔB179 above.

Construction of single gene knockout mutants

Fragments internal to the following genes were PCR amplified and introduced into *Spe*I digested pTH1937 (Milunovic *et al.* 2014): *bdhA* (oligo: DF129/DF130), *xdhA2* (DF131/DF132), *xdhB2* (DF133/DF134), *smb20847* (DF135/DF136), *smb20848* (DF137/DF138), *guaD2* (DF139/DF140), *lldD1* (DF141/DF142), and *smb20851* (DF143/DF144). Plasmids were conjugated to *S. meliloti* RmP3757, which contains the *smc03121::lacZ* gene fusion and the essential *engA* and tRNA^arg pSymB genes integrated into the chromosome. To construct the *bdhA phbB* double mutant, the *bdhA*::pTH1937 allele was first transduced into the SmFL4407 (*smc03121::lacZ*) fusion library strain followed by transduction of the *phbB*::ΩSp^R allele from Rm11347 (Aneja *et al.* 2004).

Data availability

All strains are available upon request. Gene expression data are available at GEO with the accession number GSE106129. File S1 contains detailed descriptions of all supplemental files. File S2 contains the processed RNA-seq data for chromosomal genes. File S3 contains the processed RNA-seq data for pSymB genes. File S4 contains the processed RNA-seq data for pSymA genes. File S5 contains the predicted operons in the *S. meliloti* genome. Table S1 contains strain and plasmid descriptions. Table S2 contains oligonucleotide sequences. Table S3 contains gene pangenome classifications. Table S4 lists the best putative pSymA/pSymB encoded homolog of chromosomal genes. Table S5 contains KEGG, GO, and KEGG functional enrichment data for chromosomal genes. Figure S1 displays the location of pSymB deletions. Figure S2 contains growth data for the strains used in the RNA-sequencing experiment.

RESULTS

Global transcriptional consequences of pSymA and pSymB removal

As described in Table 1, the *S. meliloti* wild type strain and the ΔpSymA, ΔpSymB, and ΔpSymAB strains each contain a different set of the three DNA replicons. The presence of the chromosome is the only consistent feature of these four strains, and previous whole genome sequencing confirmed that the chromosome of these strains is highly isogenic, with only four polymorphisms detected (diCenzo *et al.* 2016b). Thus, differences in the phenotypes of these strains can be ascribed to the presence/absence of pSymA or pSymB with high confidence, and are unlikely due to chromosomal polymorphisms.

An RNA-sequencing experiment was performed to examine how the loss of one or more replicon influences the transcriptional pattern of the remaining replicons. RNA was isolated from the above four strains grown to mid-exponential phase (OD$_{600}$ ~0.7) in minimal M9-sucrose medium. In the wild type pSymA$^+$ pSymB$^+$ strain, approximately 84% of sequencing reads mapped to the chromosome, 4% to pSymA, and 12% to pSymB. Considering replicon sizes, chromosomal transcripts were ~1.5 fold over-represented in the transcriptome, while pSymB and pSymA were ~1.9 fold and ~5 fold under-represented, respectively. This highlights the varying levels of importance of each replicon to the laboratory growth of *S. meliloti*.

The removal of the pSymA megaplasmid (~1,300 genes) had little impact on the expression of the remaining transcripts (Figures 1 and 2); no chromosomal transcripts displayed a statistically significant change in expression (p-value ≤ 0.01, fold change ≥ 2), while only six genes from four transcripts were differentially regulated on pSymB (two genes up-regulated, four genes down-regulated). This was in stark contrast to the large effect observed when the pSymB chromid (~1,600 genes) was absent (Figures 1, 2). A total of 25 chromosomal ncRNAs (22 up-regulated, 3 down-regulated) and 271 chromosomal genes (157 up-regulated, 114 down-regulated) from 172 transcripts were differentially expressed in at least one of ΔpSymB or ΔpSymAB compared to wild type, accounting for nearly 8% of all chromosomal genes. An additional 58 pSymA encoded genes (44 up-regulated, 14 down-regulated, ~4.4% of pSymA genes) from 35 transcripts were differentially expressed when pSymB was absent (Figure 1). Of the 296 chromosomal genes/ncRNAs differentially regulated in either RmP3497 or RmP3496, only 18 genes and 3 ncRNAs displayed ≥ twofold change in one strain and < 1.5 fold change in the other, suggestive of little combinatorial effect from the removal of pSymA together with pSymB. Overall, these data suggested that, in the tested conditions, the evolutionarily older pSymB chromid was more transcriptionally integrated with the chromosome than was the evolutionarily younger pSymA megaplasmid.

Reporter gene fusions to 10 promoters that were differentially expressed in the absence of pSymA and/or pSymB were identified in the *S. meliloti* fusion library collection (Cowie *et al.* 2006) and introduced into each strain to validate the RNA-seq data. These ten promoters were chosen as they either represented different expression patterns, displayed large differences in expression between strains, or we considered the genes to be particularly interesting. For all 10 of the promoter fusions tested, the results of the enzyme assays were consistent with the transcriptional patterns detected in the RNA-seq dataset (Figure 3), confirming that the RNA-seq data accurately captured the transcriptional consequences of the massive genome reductions.

Inter-replicon gene flow contributes to increased cross-replicon integration

We previously reported that an inter-replicon translocation event resulted in the transfer of a 69 kilobase region (66 genes), the ETR region, from the chromosome to the pSymB chromid in a *S. meliloti* ancestor (diCenzo *et al.* 2013). We have since integrated the corresponding ETR region from *S. fredii* NGR234 into the *S. meliloti* chromosome and produced a derivative lacking pSymA and pSymB (ΔpSymAB$_{ΩNGR69}$; Table 1) (diCenzo *et al.* 2016b). To test how the loss of this "chromosomal" gene region influenced transcription, RNA-seq was used to compare the transcriptome of ΔpSymAB$_{ΩNGR69}$ with ΔpSymAB, ΔpSymB, ΔpSymA, and wild type (Figures 1, 2). Mapping the sequencing reads to the *S. fredii* NGR234 genome confirmed that the genes of the ETR region in the ΔpSymAB$_{ΩNGR69}$ genome were expressed (not shown).

A

ΔpSymAB ΔpSymAB ΔpSymB ΔpSymA
ΩNGR

B

C

ΔpSymA

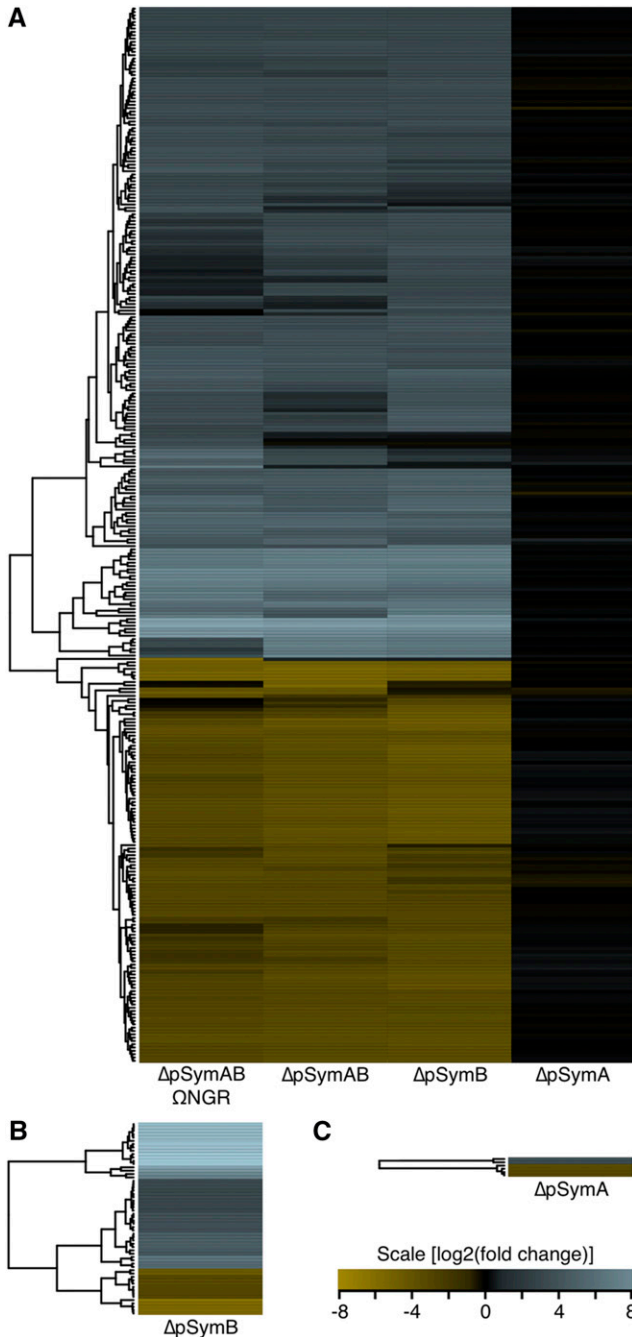Scale [log2(fold change)]

-8   -4   0   4   8

ΔpSymB

Figure 1 Overview of the transcriptional consequences of replicon removal. Heatmaps summarizing the expression of all genes differentially expressed (≥ twofold, p-value ≤ 0.01) in at least one strain are shown. (A) The 290 chromosomal genes and 25 chromosomal ncRNAs, (B) the 58 pSymA genes, (C) and the six pSymB genes differentially expressed are shown. Expression of all genes are shown relative to that of the wild type *S. meliloti* strain RmP3499 (pSymA+ pSymB+). Those in blue are up-regulated compared to wild type while those in yellow are down-regulated compared to wild type. Expression changes are displayed on a log₂ scale.

The up-regulation of seven genes from two transcripts (*smc03116-smc03121*, and *smc04246*) in the absence of pSymB could be traced primarily to the loss of the ETR region as these genes were expressed two to sixfold higher in ΔpSymAB than in ΔpSymAB_ΩNGR69 (File S2;
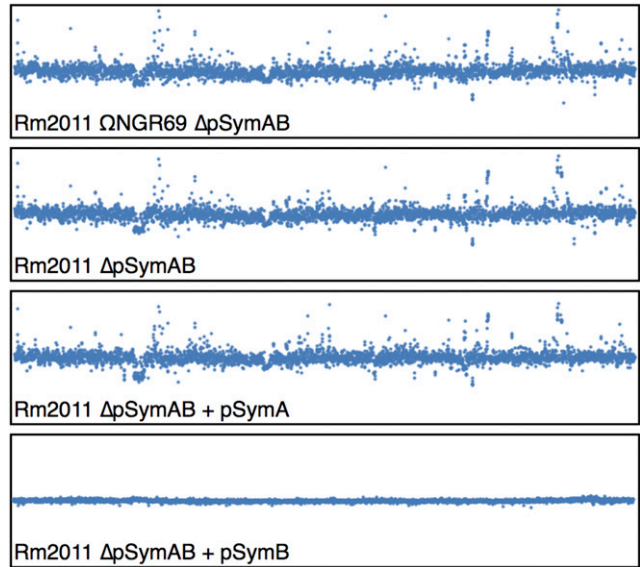
Figure 2 Chromosomal localization of the differentially expressed genes. Each box displays the expression of all 3824 genes and ncRNAs on the *S. meliloti* chromosome, with the genes displayed in the same order they are situated on the chromosome. Each dot represents a single gene, and displays the expression of the gene relative to that of the wild type *S. meliloti* strain RmP3499 (pSymA+ pSymB+). Expression changes are displayed on a log₂ scale, with the top and bottom of each box representing a 64-fold up-regulation or down-regulation relative to wild type.

see *smc03121* in Figure 3 for an example). Several additional genes displayed an intermediate phenotype in ΔpSymAB_ΩNGR69 compared to ΔpSymAB and the wild type, suggesting the phenotype was partially linked to the loss of the ETR region. Eight genes from five transcripts were up-regulated compared to wild type primarily when the ETR region was present but the rest of pSymB was absent; these genes were expressed ∼ two to ninefold higher in ΔpSymAB_ΩNGR69 than in ΔpSymAB. Finally, one gene (*smc03107*) was down-regulated compared to wild type only in the absence of pSymB if the ETR region was present. These data revealed that the flow of genes between replicons in the *S. meliloti* genome contributed to their transcriptional integration. Additionally, these data illustrated how genomic context influenced the transcriptional consequences of gene deletion/addition.

Causes of transcriptional changes can be traced to specific regions of pSymB

It was possible that the observed transcriptional changes were a consequence of the reduced growth rate of strains lacking pSymB instead of a direct consequence of loss of specific pSymB encoded genes; in the tested medium, the generation time of strains lacking pSymB was ∼5.3 hr compared to ∼3.1 hr for the wild type (Figure S2). Therefore, expression from four promoters up-regulated in the absence of pSymB was examined in a collection of 11 large-scale pSymB deletion mutants that cumulatively remove ∼98% of pSymB (Figure S1) (Milunovic *et al.* 2014); these deletion mutants displayed little growth rate difference compared to the wild type (see well D11 of plate 1 of the Biolog Phenotype MicroArray data of diCenzo *et al.* 2016a). Expression from the divergently transcribed promoters of the five gene *sma2085-sma2077* and eight gene *sma2087-sma2101* operons on pSymA were observed to be up-regulated specifically in the ΔB180 (pSymB nt. 635,940-869,642) deletion (Figure 4). Expression of the chromosomal
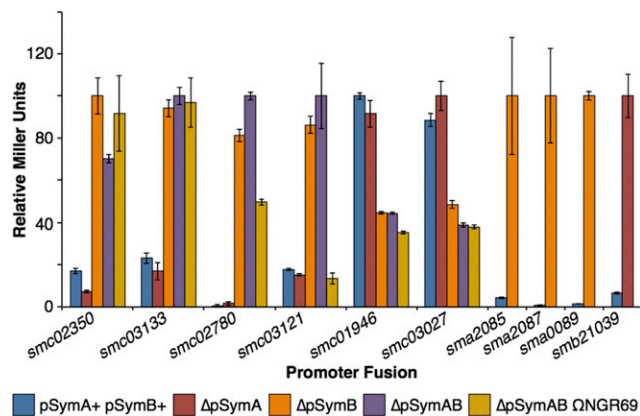
Figure 3 Reporter gene fusion confirmation of the RNA-seq data. Transcriptional fusions of genes of interest to either a *lacZ* or *gusA* reporter gene were prepared in each strain used for the RNA-seq analysis. The β-galactosidase or β-glucuronidase activities for each fusion, expressed as relative Miller Units, are shown. For each fusion, the values are given relative to the strain with the highest activity, and the strain with the highest activity was normalized to a value of 100. Data points represent the means of triplicate samples, with the error bars indicating the standard error.
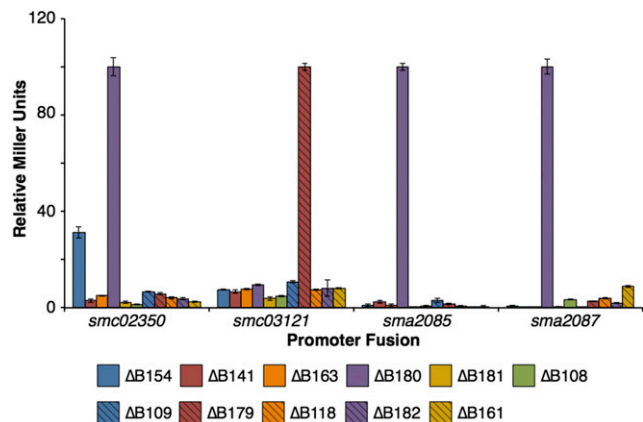

Figure 4 Localization of pSymB regions that influence expression of chromosome or pSymA genes. Transcriptional *lacZ* or *gusA* reporter gene fusions to genes of interest recombined into the same strain as each of the eleven large-scale pSymB deletion that cumulatively span ~98% of pSymB. The β-galactosidase or β-glucuronidase activities for each fusion, expressed as relative Miller Units, are shown. For each fusion, the values are given relative to the strain with the highest activity, and the strain with the highest activity was normalized to a value of 100. The location of the deletions are shown in Figure S1. Data points represent the means of triplicate samples, with the error bars indicating the standard error.

promoter of the six gene *smc0235-smc02344* operon was up-regulated in both the ΔB180 and ΔB154 (pSymB nt. 62,137-100,636) deletions (Figure 4). Finally, the six gene *smc03121-smc03116* operon was specifically up-regulated in the ΔB179 (pSymB nt. 1,207,052-1,322,226) deletion (Figure 4). Hence, while the expression of an unknown number of genes may have simply been a consequence of differences in growth rates, these observations confirmed that several of the transcriptional changes were directly attributable to the loss of specific pSymB encoded genes.

Few detectable biases in chromosomal genes up-regulated in the absence of pSymB

We attempted to determine whether there was any feature that separated chromosomal genes up-regulated in the absence of pSymB from the average chromosomal gene. As can be seen in Figure 2, genes up-regulated upon removal of pSymB were distributed across the entire chromosome with no obvious locational bias. In order to limit the number of false positives, only the 146 genes up-regulated (≥twofold) in at least two of the three strains lacking pSymB were considered for the subsequent analyses. No statistically significant biases were observed in the pangenome classification (core, accessory, unique) of genes up-regulated in the absence of pSymB compared to genes whose expression was unchanged or down-regulated (Table S3).

Greater than 10% of *S. meliloti* chromosomal genes may have a functionally redundant copy on either pSymA or pSymB (diCenzo and Finan 2015), and we therefore wondered if the up-regulation was a result of loss of pSymB encoded homologs. A Blast bidirectional best hit approach was used to predict putative functional homologs between chromosomal genes and genes on pSymA/pSymB (see Materials and Methods). It was observed that of the genes whose expression did not change across strains, ~5% had a best putative functional homolog on pSymA and an additional ~5% had a best putative functional homolog on pSymB (Table S4). Of the genes up-regulated in the absence of pSymB, ~5% had a best putative functional homolog on pSymA, whereas ~15% had a best putative functional homolog on pSymB (Table S4). This therefore supported that some of genes

up-regulated in the absence of pSymB may have been an attempt to compensate for the loss of a functional homolog, although the majority of changes could not be explained by functional redundancy.

Genes were annotated with COG terms, GO terms, and KEGG pathway terms in an attempt to identify functional biases between genes up-regulated or not up-regulated upon loss of pSymB (Table S5). The only biases observed were in the broad COG categories of inorganic ion transport and metabolism (COG P), transcription (COG K), and energy production and conversion (COG C). No biases were observed in GO term or KEGG pathway annotations. Hence, no obvious functional role separated the genes up-regulated when pSymB was removed compared to the average chromosomal gene.

Cell motility is decreased in the absence of pSymB

Genes down-regulated in the absence of pSymB are spread across the chromosome; however, a clear cluster of down-regulated genes was evident (Figure 2). This cluster contained flagellum related genes. When a functional annotation (COG, GO, and KEGG pathways) of the 90 genes down-regulated (≤ twofold) in at least two strains lacking pSymB was performed, a clear enrichment was observed in sensory transduction, motility, and chemotaxis related processes (Table S5). Down-regulation of the ten gene *flgB-fliP* operon was confirmed using a transcriptional gene fusions (fusion *smc03027* in Figure 3). These results therefore indicated that the loss of pSymB results in decreased expression of taxis related genes in the tested medium. Indeed, in earlier experiments we observed that strains lacking pSymB displayed reduced swimming motility in TY soft agar plates (data not shown), although we could not rule out that this was simply a consequence of the reduced growth rate.

Loss of bdhA results in smc03121 up-regulation through 3-hydroxybutyrate accumulation

The data presented earlier indicated that the *smc03121* operon was transcriptionally up-regulated in the ΔpSymB and ΔpSymAB strains,

that this phenotype could be traced to the B179 region of pSymB that includes the ETR region (Figure 4), and that re-introduction of just the ETR region was sufficient to restore wild type levels of smc03121 expression (Figure 3). Using a series of sub-deletions within the B179 region, the responsible locus within the B179 region was narrowed down to a 12 kb region (pSymB nt. 1,242,846-1,255,032) spanning eight complete genes (data not shown; Figure 5A). Single cross-over plasmid integration was used to construct knockouts of seven of these eight genes in a wild type background containing a smc03121::lacZ fusion (Figure 5A), and it was observed that disrupting just the bdhA gene, encoding a 3-hydroxybutyrate dehydrogenase (Aneja and Charles 1999), resulted in transcriptional up-regulation of smc03121 (Figure 5B). This result was also seen in a recent RNA-seq study of *S. meliloti* poly-3-hydroxybutyrate (PHB) cycle mutants (D'Alessio *et al.* 2017). However, smc03121 was induced and expressed equally in all eight gene mutants when the medium was supplemented with 10 mM 3-hydroxybutyrate (Figure 5B). We therefore reasoned that 3-hydroxybutyrate may accumulate, and be secreted into the medium, in a bdhA null mutant, resulting in the observed smc03121 up-regulation. As the sole source of 3-hydroxybutyrate in the *S. meliloti* iGD1575 *in silico* metabolic reconstruction is from PHB breakdown (diCenzo *et al.* 2016a), a phbB::ΩSp mutation was introduced into a smc03121::lacZ bdhA::pTH1937 strain as this mutation prevents PHB synthesis (Aneja *et al.* 2004). Indeed, expression of smc03121 was lower in the bdhA phbB double mutant than in the bdhA single mutant when grown in M9-sucrose minimal medium lacking 3-hydroxybutyrate (Figure 5B). Hence, it was concluded that, like in *Ralstonia eutrophus* (Doi *et al.* 1990; Taidi *et al.* 1995), continual PHB synthesis and degradation was occurring during exponential growth in M9-sucrose, and that mutation of bdhA resulted in the accumulation and export of the BdhA substrate, 3-hydroxybutyrate, thereby resulting in the induction of the smc03121-smc03116 transport and metabolic operon.

DISCUSSION

The *S. meliloti* genome is a good model for the study of the multipartite genome as it contains three replicons each with distinct characteristics. In particular, the pSymA megaplasmid is derived primarily through recent and ongoing HGT whereas pSymB is a more stable replicon derived largely through ancient HGT (Galardini *et al.* 2013). Here, RNA-sequencing was used to obtain a global overview of the transcriptional consequences of the removal of these large secondary replicons from the genome to gain insights into the level of transcriptional integration of these replicons and the influence of the age of the replicon. Importantly, it was possible to link transcriptional changes to the loss of specific gene regions. This confirmed that several of the observed expression changes are not simply indirect consequences of the reduced growth rate, but instead reflect regulatory or metabolic interactions between the replicons.

The absence of pSymB resulted in major transcriptome changes with ~8% of chromosomal genes and ~4% of pSymA genes differing in expression (Figure 1). This was consistent with metabolomics work that identified large perturbations in the metabolome of strains lacking pSymB (Fei *et al.* 2016). In contrast, the data reported here indicated that removal of pSymA impacted expression of a mere four operons (Figure 1). This was similar to previous 2D gel electrophoretic work that detected only 30 chromosome or pSymB encoded proteins that differed in abundance following pSymA removal (Chen *et al.* 2000), and metabolomics work that observed few metabolic differences between strains with and without pSymA (Fei *et al.* 2016). However, transcriptional integration of pSymA with the chromosome may still occur during specialized processes; for example, the pSymA-encoded
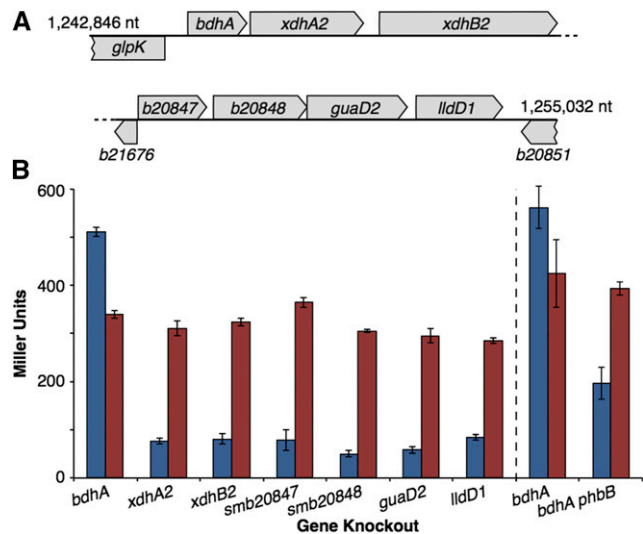
Figure 5 Effect of gene knockouts and 3-hydroxybutyrate on smc03131 expression. (A) Schematic of the gene organization of the 12 kb region whose deletion resulted in induction of smc03121. (B) β-Galactosidase activity from a smc03121::lacZ fusion in either M9-sucrose minimal medium (blue) or M9-sucrose supplemented with 10 mM 3-hydroxybutyrate (red). The gene(s) mutated in each strain are indicated along the x-axis, and expression values are presented as Miller units. The vertical dashed line separates assays that were performed in independent experiments. Values represent the mean of triplicate samples, with the error bars indicating the standard error.

symbiotic transcriptional regulator FixJ has chromosomally encoded target genes (Bobik *et al.* 2006). Overall, these studies support that newly acquired replicons are poorly integrated into the central transcriptional and metabolic networks of the cell, and that continual co-existence of two replicons results in a gradual inter-dependency of the replicons at multiple levels.

The analyses described here allowed an examination of how secondary replicon removal influences chromosomal gene expression, but they did not allow study of how transcription of secondary replicons is influenced by the primary replicon. However, the effects of pSymB removal *vs.* pSymA removal on the transcription of genes on the other replicon can serve as a proxy since pSymB has several chromosome-like features and was already present at the time pSymA was acquired. Whereas the loss of pSymB influenced expression of ~4% of pSymA genes, less than 0.5% of pSymB genes were impacted by the loss of pSymA (Figure 1). Previously, an *in silico* analysis predicted that approximately twofold more genes on pSymA or pSymB are regulated by chromosomal transcription factors than chromosomal genes regulated by pSymA or pSymB encoded transcription factors (Galardini *et al.* 2015). Together, these results suggest that existing replicons 'domesticate' newly acquired replicons; *i.e.*, there is a higher likelihood of genes on newly acquired replicons being integrated into existing regulatory networks than there is of existing genes being transcriptionally modified in response to acquisition of the new replicon.

Plasmids and other secondary replicons often have specialized functions, and have been postulated to play a key role in adaptation and evolution (Reanney 1976). A transcriptional separation from chromosomal genes may help facilitate successful horizontal transfer of the replicon by reducing the occurrence of off-target effects. Selection may therefore have favored the spread of replicons lacking significant ties to chromosomal elements, potentially explaining, in part, the low level of integration of pSymA into the central cellular networks. Thus the level

of integration of two replicons in a genome may be a function of selection for improved horizontal transfer efficiency and the length of replicon co-evolution. However, whether or not increased integration in central networks only occurs after the replicon has largely lost the ability to undergo horizontal transfer requires further study.

The transcriptional inter-dependence of each replicon in a genome can be mediated through multiple mechanisms. The simplest explanation is that genes are integrated into the regulons of transcription factors located on other replicons as has been predicted *in silico* (Galardini *et al.* 2015). Another possibility is through the transfer of genes between replicons. It was shown that the transcriptional changes of seven chromosomal genes following the removal of pSymB were dependent on the loss of the ETR region that translocated to pSymB in a recent *S. meliloti* ancestor (Figures 1, 3) (diCenzo *et al.* 2013; diCenzo *et al.* 2016b). Functional redundancies can have an influence as well. Maintenance of functional redundancy can be driven by reductions in expression of the duplicate genes (Qian *et al.* 2010), and deletion of one of the gene could therefore be complemented by induction of the other. The data presented here supports this notion as genes with putative homologs on pSymB are over-represented in the gene set that is up-regulated following pSymB removal. These include several ABC-type transport systems and cytochrome proteins (Table S4, File S2). Moreover, the transcriptional changes could be indirect responses mediated through shifts in the metabolic profile of the cell or environment. This was exemplified by the *smc03131* operon, whose induction by the removal of pSymB appeared to be a consequence of an accumulation of 3-hydroxybutyrate due to an inability of strains lacking pSymB to degrade the 3-hydroxybutyrate produced through PHB breakdown (Figure 5).

Considering that pSymA and pSymB were formed largely through either recent or ancient HGT, respectively, the data presented here can be interpreted more broadly with respect to HGT. In particular, this data may suggest that a gene acquired through recent HGT is more likely to be maintained in the genome if it does not immediately disturb the existing central transcriptional network, with transcriptional integration being a gradual process that occurs following the gene's acquisition. In support of this, transcriptional regulators acquired through HGT primarily regulate genes that are co-transferred with the regulator whereas global regulators are primarily vertically transmitted (Price *et al.* 2008). Previous work has also indicated that recently acquired genes are regulated by few transcription factors and that it takes millions of years for regulation to reach levels similar to that of native genes (Lercher and Pál 2008). Moreover, horizontally acquired genes often encode functions located at the periphery of metabolic networks (Pál, Papp, and Lercher 2005a), and genes most likely to be maintained following their acquisition generally have low rates of transcription or translation (Taoka *et al.* 2004; Park and Zhang 2012).

We produced the genome reduced strains used here in order to study the genome biology and evolution of *S. meliloti* as well as the contributions of these replicons to nitrogen fixing symbiosis (diCenzo *et al.* 2014; diCenzo *et al.* 2016b). This work completes our systems-level analyses of these strains, which has included phenomics (diCenzo *et al.* 2014), metabolomics (Fei *et al.* 2016), and now transcriptomics. The ΔpSymAB strain represents a genome reduction of nearly 2,900 genes and displays many phenotypic, metabolomic, and transcriptomic perturbations. As shown by the work with *smc03131* and *bhdA* in this study (Figure 5), these data sets together with the existing *S. meliloti* genomic resources, including the fusion and deletion libraries (Cowie *et al.* 2006; Milunovic *et al.* 2014), can facilitate rapid linkage of phenotypes to the loss of specific genes. As such, we expect future mining of these data to reveal many interesting insights into the genetics of *S. meliloti*.

LITERATURE CITED

Amorós-Moya, D., S. Bedhomme, M. Hermann, and I. G. Bravo, 2010 Evolution in regulatory regions rapidly compensates the cost of nonoptimal codon usage. Mol. Biol. Evol. 27(9): 2141–2151. https://doi.org/10.1093/molbev/msq103

Anders, S., P. T. Pyl, and W. Huber, 2015 HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31(2): 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, S., 2010—FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aneja, P., and T. C. Charles, 1999 Poly-3-hydroxybutyrate degradation in *Rhizobium* (*Sinorhizobium*) *meliloti*: isolation and characterization of a gene encoding 3-hydroxybutyrate dehydrogenase. J. Bacteriol. 181: 849–857.

Aneja, P., M. Dai, D. A. Lacorre, B. Pillon, and T. C. Charles, 2004 Heterologous complementation of the exopolysaccharide synthesis and carbon utilization phenotypes of *Sinorhizobium meliloti* Rm1021 polyhydroxyalkanoate synthesis mutants. FEMS Microbiol. Lett. 239(2): 277–283. https://doi.org/10.1016/j.femsle.2004.08.045

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler *et al.*, 2000 Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25(1): 25–29. https://doi.org/10.1038/75556

Baek, J. H., and D. K. Chattoraj, 2014 Chromosome I controls chromosome II replication in *Vibrio cholerae*. PLoS Genet. 10(2): e1004184. https://doi.org/10.1371/journal.pgen.1004184

Baltrus, D. A., 2013 Exploring the costs of horizontal gene transfer. Trends Ecol. Evol. 28(8): 489–495. https://doi.org/10.1016/j.tree.2013.04.002

Barnett, M. J., C. J. Toman, R. F. Fisher, and S. R. Long, 2004 A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote–host interaction. Proc. Natl. Acad. Sci. USA 101(47): 16636–16641. https://doi.org/10.1073/pnas.0407269101

Baumdicker, F., W. R. Hess, and P. Pfaffelhuber, 2012 The infinitely many genes model for the distributed genome of bacteria. Genome Biol. Evol. 4(4): 443–456. https://doi.org/10.1093/gbe/evs016

Becker, A., H. Bergès, E. Krol, C. Bruand, S. Rüberg *et al.*, 2004 Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. Mol. Plant Microbe Interact. 17(3): 292–303. https://doi.org/10.1094/MPMI.2004.17.3.292

Bobik, C., E. Meilhoc, and J. Batut, 2006 FixJ: a major regulator of the oxygen limitation response and late symbiotic functions of *Sinorhizobium meliloti*. J. Bacteriol. 188(13): 4890–4902. https://doi.org/10.1128/JB.00251-06

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10(1): 421. https://doi.org/10.1186/1471-2105-10-421

Chen, H., J. Higgins, I. J. Oresnik, M. F. Hynes, S. Natera *et al.*, 2000 Proteome analysis demonstrates complex replicon and luteolin interactions in pSyma-cured derivatives of *Sinorhizobium meliloti* strain 2011. Electrophoresis 21(17): 3833–3842. https://doi.org/10.1002/1522-2683(200011)21:17<3833::AID-ELPS3833>3.0.CO;2-I

Cheng, J., B. Poduska, R. A. Morton, and T. M. Finan, 2011 An ABC-type cobalt transport system is essential for growth of *Sinorhizobium meliloti* at trace metal concentrations. J. Bacteriol. 193(17): 4405–4416. https://doi.org/10.1128/JB.05045-11

Cohen, O., U. Gophna, and T. Pupko, 2011 The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. Mol. Biol. Evol. 28(4): 1481–1489. https://doi.org/10.1093/molbev/msq333

Cooper, V. S., S. H. Vohr, S. C. Wrocklage, and P. J. Hatcher, 2010 Why genes evolve faster on secondary chromosomes in bacteria. PLOS Comput. Biol. 6(4): e1000732. https://doi.org/10.1371/journal.pcbi.1000732

Cowie, A., J. Cheng, C. D. Sibley, Y. Fong, R. Zaheer *et al.*, 2006   An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. Appl. Environ. Microbiol. 72(11): 7156–7167. https://doi.org/10.1128/AEM.01397-06

D'Alessio, M., R. Nordeste, A. C. Doxey, and T. C. Charles, 2017   Transcriptome analysis of polyhydroxybutyrate cycle mutants reveals discrete loci connecting nitrogen utilization and carbon storage in Sinorhizobium meliloti. mSystems 2 e00035–17. https://doi.org/10.1128/mSystems.00035-17

diCenzo, G., B. Milunovic, J. Cheng, and T. M. Finan, 2013   The tRNA^{arg} gene and *engA* are essential genes on the 1.7-mb pSymB megaplasmid of *Sinorhizobium meliloti* and were translocated together from the chromosome in an ancestral strain. J. Bacteriol. 195(2): 202–212. https://doi.org/10.1128/JB.01758-12

diCenzo, G. C., and T. M. Finan, 2015   Genetic redundancy is prevalent within the 6.7 Mb *Sinorhizobium meliloti* genome. Mol. Genet. Genomics 290(4): 1345–1356. https://doi.org/10.1007/s00438-015-0998-6

diCenzo, G. C., and T. M. Finan, 2017   The divided bacterial genome: structure, function, and evolution. Microbiol. Mol. Biol. Rev. 81: e00019–17–17. https://doi.org/10.1128/MMBR.00019-17

diCenzo, G. C., A. Checcucci, M. Bazzicalupo, A. Mengoni, C. Viti *et al.*, 2016a   Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. Nat. Commun. 7: 12219. https://doi.org/10.1038/ncomms12219

diCenzo, G. C., A. M. MacLean, B. Milunovic, G. B. Golding, and T. M. Finan, 2014   Examination of prokaryotic multipartite genome evolution through experimental genome reduction. PLoS Genet. 10(10): e1004742. https://doi.org/10.1371/journal.pgen.1004742

diCenzo, G. C., Z. Muhammed, M. Østerås, S. A. P. O'Brien, and T. M. Finan, 2017   A key regulator of the glycolytic and gluconeogenic central metabolic pathways in *Sinorhizobium meliloti*. Genetics 207(3): 961–974. https://doi.org/10.1534/genetics.117.300212

diCenzo, G. C., M. Zamani, B. Milunovic, and T. M. Finan, 2016b   Genomic resources for identification of the minimal N$_2$-fixing symbiotic genome. Environ. Microbiol. 18(8): 2534–2547. https://doi.org/10.1111/1462-2920.13221

Doi, Y., A. Segawa, Y. Kawaguchi, and M. Kunioka, 1990   Cyclic nature of poly(3-hydroxyalkanoate) metabolism in *Alcaligenes eutrophus*. FEMS Microbiol. Lett. 55(1-2): 165–169. https://doi.org/10.1111/j.1574-6968.1990.tb13856.x

Fei, F., G. C. diCenzo, D. M. E. Bowdish, B. E. McCarry, and T. M. Finan, 2016   Effects of synthetic large-scale genome reduction on metabolism and metabolic preferences in a nutritionally complex environment. Metabolomics 12(2): 23. https://doi.org/10.1007/s11306-015-0928-y

Finan, T. M., E. Hartweig, K. LeMieux, K. Bergman, G. C. Walker *et al.*, 1984   General transduction in *Rhizobium meliloti*. J. Bacteriol. 159: 120–124.

Galardini, M., M. Brilli, G. Spini, M. Rossi, B. Roncaglia *et al.*, 2015   Evolution of intra-specific regulatory networks in a multipartite bacterial genome. PLOS Comput. Biol. 11(9): e1004478. https://doi.org/10.1371/journal.pcbi.1004478

Galardini, M., F. Pini, M. Bazzicalupo, E. G. Biondi, and A. Mengoni, 2013   Replicon-dependent bacterial genome evolution: the case of *Sinorhizobium meliloti*. Genome Biol. Evol. 5(3): 542–558. https://doi.org/10.1093/gbe/evt027

Galibert, F., T. M. Finan, S. R. Long, A. Pühler, A. P. Abola *et al.*, 2001   The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293(5530): 668–672. https://doi.org/10.1126/science.1060966

Galli, E., M. Poidevin, R. Le Bars, J.-M. Desfontaines, L. Muresan *et al.*, 2016   Cell division licensing in the multi-chromosomal *Vibrio cholerae* bacterium. Nat. Microbiol. 1(9): 16094. https://doi.org/10.1038/nmicrobiol.2016.94

González, V., R. I. Santamaría, P. Bustos, I. Hernández-González, A. Medrano-Soto *et al.*, 2006   The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc. Natl. Acad. Sci. USA 103(10): 3834–3839. https://doi.org/10.1073/pnas.0508502103

Guo, H., S. Sun, B. Eardly, T. Finan, and J. Xu, 2009   Genome variation in the symbiotic nitrogen-fixing bacterium *Sinorhizobium meliloti*. Genome 52(10): 862–875. https://doi.org/10.1139/G09-060

Hao, W., and G. B. Golding, 2006   The fate of laterally transferred genes: life in the fast lane to adaptation or death. Genome Res. 16(5): 636–643. https://doi.org/10.1101/gr.4746406

Harrison, P. W., R. P. J. Lower, N. K. D. Kim, and J. P. W. Young, 2010   Introducing the bacterial "chromid": not a chromosome, not a plasmid. Trends Microbiol. 18(4): 141–148. https://doi.org/10.1016/j.tim.2009.12.010

Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn *et al.*, 2000   DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406(6795): 477–483. https://doi.org/10.1038/35020000

Huerta-Cepas, J., D. Szklarczyk, K. Forslund, H. Cook, D. Heller *et al.*, 2016   eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44(D1): D286–D293. https://doi.org/10.1093/nar/gkv1248

Jeong, J.-Y., H.-S. Yim, J.-Y. Ryu, H. S. Lee, J.-H. Lee *et al.*, 2012   One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. Appl. Environ. Microbiol. 78(15): 5440–5443. https://doi.org/10.1128/AEM.00844-12

Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, 2016   KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44(D1): D457–D462. https://doi.org/10.1093/nar/gkv1070

Lawrence, J. G., and H. Ochman, 1998   Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA 95(16): 9413–9417. https://doi.org/10.1073/pnas.95.16.9413

Lercher, M. J., and C. Pál, 2008   Integration of horizontally transferred genes into regulatory interaction networks takes many million years. Mol. Biol. Evol. 25(3): 559–567. https://doi.org/10.1093/molbev/msm283

Li, H., and R. Durbin, 2009   Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14): 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Love, M. I., W. Huber, and S. Anders, 2014   Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12): 550. https://doi.org/10.1186/s13059-014-0550-8

López-Guerrero, M. G., E. Ormeño-Orrillo, J. L. Acosta, A. Mendoza-Vargas, M. A. Rogel *et al.*, 2012   Rhizobial extrachromosomal replicon variability, stability and expression in natural niches. Plasmid 68(3): 149–158. https://doi.org/10.1016/j.plasmid.2012.07.002

Martin, M., 2011   Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: pp. 10–12. https://doi.org/10.14806/ej.17.1.200

Milunovic, B., G. C. diCenzo, R. A. Morton, and T. M. Finan, 2014   Cell growth inhibition upon deletion of four toxin-antitoxin loci from the megaplasmids of *Sinorhizobium meliloti*. J. Bacteriol. 196(4): 811–824. https://doi.org/10.1128/JB.01104-13

Ochman, H., J. G. Lawrence, and E. A. Groisman, 2000   Lateral gene transfer and the nature of bacterial innovation. Nature 405(6784): 299–304. https://doi.org/10.1038/35012500

Oresnik, I. J., S. L. Liu, C. K. Yost, and M. F. Hynes, 2000   Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability. J. Bacteriol. 182(12): 3582–3586. https://doi.org/10.1128/JB.182.12.3582-3586.2000

Park, C., and J. Zhang, 2012   High expression hampers horizontal gene transfer. Genome Biol. Evol. 4(4): 523–532. https://doi.org/10.1093/gbe/evs030

Pál, C., B. Papp, and M. J. Lercher, 2005a   Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet. 37(12): 1372–1375. https://doi.org/10.1038/ng1686

Pál, C., B. Papp, and M. J. Lercher, 2005b   Horizontal gene transfer depends on gene content of the host. Bioinformatics 21(Suppl 2): ii222–ii223. https://doi.org/10.1093/bioinformatics/bti1136

Pini, F., N. J. De Nisco, L. Ferri, J. Penterman, A. Fioravanti *et al.*, 2015 Cell cycle control by the master regulator CtrA in *Sinorhizobium meliloti*. PLoS Genet. 11(5): e1005232. https://doi.org/10.1371/journal.pgen.1005232

Price, M. N., P. S. Dehal, and A. P. Arkin, 2008 Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. Genome Biol. 9(1): R4. https://doi.org/10.1186/gb-2008-9-1-r4

Qian, W., B.-Y. Liao, A. Y.-F. Chang, and J. Zhang, 2010 Maintenance of duplicate genes and their functional redundancy by reduced expression. Trends Genet. 26(10): 425–430. https://doi.org/10.1016/j.tig.2010.07.002

Ramachandran, V. K., A. K. East, R. Karunakaran, J. A. Downie, and P. S. Poole, 2011 Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. Genome Biol. 12(10): R106. https://doi.org/10.1186/gb-2011-12-10-r106

Reanney, D., 1976 Extrachromosomal elements as possible agents of adaptation and development. Bacteriol. Rev. 40: 552–590.

Sallet, E., B. Roux, L. Sauviac, M.-F. Jardinaud, S. Carrère *et al.*, 2013 Next-generation annotation of prokaryotic genomes with EuGene-P: application to *Sinorhizobium meliloti* 2011. DNA Res. 20(4): 339–354. https://doi.org/10.1093/dnares/dst014

Sambrook, J., E. F. Fritsch, and T. Maniatis, 1989 Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, New York.

San Millan, A., M. Toll-Riera, Q. Qi, and R. C. MacLean, 2015 Interactions between horizontally acquired genes create a fitness cost in *Pseudomonas aeruginosa*. Nat. Commun. 6(1): 6845. https://doi.org/10.1038/ncomms7845

Sorek, R., Y. Zhu, C. J. Creevey, M. P. Francino, P. Bork *et al.*, 2007 Genome-wide experimental determination of barriers to horizontal gene transfer. Science 318(5855): 1449–1452. https://doi.org/10.1126/science.1147112

Sugawara, M., B. Epstein, B. D. Badgley, T. Unno, L. Xu *et al.*, 2013 Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. Genome Biol. 14(2): R17. https://doi.org/10.1186/gb-2013-14-2-r17

Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc, 2011 REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 6(7): e21800. https://doi.org/10.1371/journal.pone.0021800

Taidi, B., D. A. Mansfield, and A. J. Anderson, 1995 Turnover of poly(3-hydroxybutyrate) (PHB) and its influence on the molecular mass of the polymer accumulated by *Alcaligenes eutrophus* during batch culture. FEMS Microbiol. Lett. 129: 201–205.

Taoka, M., Y. Yamauchi, T. Shinkawa, H. Kaji, W. Motohashi *et al.*, 2004 Only a small subset of the horizontally transferred chromosomal genes in *Escherichia coli* are translated into proteins. Mol. Cell. Proteomics 3(8): 780–787. https://doi.org/10.1074/mcp.M400030-MCP200

Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin *et al.*, 2003 The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4(1): 41. https://doi.org/10.1186/1471-2105-4-41

Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini *et al.*, 2005 Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. USA 102(39): 13950–13955 (erratum: Proc. Natl. Acad. Sci. USA 102: 16530). https://doi.org/10.1073/pnas.0506758102

Tjaden, B., 2015 *De novo* assembly of bacterial transcriptomes from RNA-seq data. Genome Biol. 16(1): 1. https://doi.org/10.1186/s13059-014-0572-2

Val, M.-E., M. Marbouty, F. de Lemos Martins, S. P. Kennedy, H. Kemble *et al.*, 2016 A checkpoint control orchestrates the replication of the two chromosomes of *Vibrio cholerae*. Sci. Adv. 2(4): e1501914. https://doi.org/10.1126/sciadv.1501914

Warnes, G. R., B. Bolker, L. Bonebakker, R. Gentleman, W. Huber *et al.*, 2009 gplots: Various R programming tools for plotting data. R package version 2. http://CRAN.R-project.org/package=gplots

Wong, K., and G. B. Golding, 2003 A phylogenetic analysis of the pSymB replicon from the *Sinorhizobium meliloti* genome reveals a complex evolutionary history. Can. J. Microbiol. 49(4): 269–280. https://doi.org/10.1139/w03-037

Xu, Q., M. Dziejman, and J. J. Mekalanos, 2003 Determination of the transcriptome of *Vibrio cholerae* during intraintestinal growth and mid-exponential phase *in vitro*. Proc. Natl. Acad. Sci. USA 100(3): 1286–1291. https://doi.org/10.1073/pnas.0337479100

Yoder-Himes, D. R., K. T. Konstantinidis, and J. M. Tiedje, 2010 Identification of potential therapeutic targets for *Burkholderia cenocepacia* by comparative transcriptomics. PLoS One 5(1): e8724. https://doi.org/10.1371/journal.pone.0008724

Zhang, Y., T. Aono, P. Poole, and T. M. Finan, 2012 NAD(P)+-malic enzyme mutants of *Sinorhizobium* sp. strain NGR234, but not *Azorhizobium caulinodans* ORS571, maintain symbiotic N_2 fixation capabilities. Appl. Environ. Microbiol. 78(8): 2803–2812. https://doi.org/10.1128/AEM.06412-11

Communicating editor: R. Doerge