Mol Genet Genomics
DOI 10.1007/s00438-015-0998-6

ORIGINAL PAPER

Genetic redundancy is prevalent within the 6.7 Mb *Sinorhizobium meliloti* genome

George C. diCenzo · Turlough M. Finan

Received: 4 December 2014 / Accepted: 17 January 2015

Abstract Biological pathways are frequently identified via a genetic loss-of-function approach. While this approach has proven to be powerful, it is imperfect as illustrated by well-studied pathways continuing to have missing steps. One potential limiting factor is the masking of phenotypes through genetic redundancy. The prevalence of genetic redundancy in bacterial species has received little attention, although isolated examples of functionally redundant gene pairs exist. Here, we made use of a strain of *Sinorhizobium meliloti* whose genome was reduced by 45 % through the complete removal of a megaplasmid and a chromid (3 Mb of the 6.7 Mb genome was removed) to begin quantifying the level of genetic redundancy within a large bacterial genome. A mutagenesis of the strain with the reduced genome identified a set of transposon insertions precluding growth of this strain on minimal medium. Transfer of these mutations to the wild-type background revealed that 10–15 % of these chromosomal mutations were located within duplicated genes, as they did not prevent growth of cells with the full genome. The functionally redundant genes were involved in a variety of metabolic pathways, including central carbon metabolism, transport, and amino acid biosynthesis. These results indicate that genetic redundancy may be prevalent within large bacterial genomes. Failing to account for redundantly encoded functions in loss-of-function studies will impair our understanding of a broad range of biological processes and limit our ability to use synthetic biology in the construction of designer cell factories.

Keywords Genomics · Evolution · Minimal · Multipartite · Arginine · Proline

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-015-0998-6) contains supplementary material, which is available to authorized users.

G. C. diCenzo · T. M. Finan (✉)
Department of Biology, McMaster University, 1280 Main St. W., Hamilton, ON L8S 4K1, Canada
e-mail: finan@mcmaster.ca

Introduction

Despite significant advances in the annotation of prokaryotic genomes, a surprisingly large percentage of genes remain uncharacterized. As of October 17, 2014, a search of the Entrez Gene database (Maglott et al. 2011) with the query 'hypothetical' returns ~32 % of all prokaryotic genes (~2,900,000 of ~8,900,000 entries), of which ~1,750,000 are returned with a query of 'conserved hypothetical'. Whereas some hypothetical genes may represent artefacts of automated genome annotation, many of the conserved hypothetical genes are likely to encode true proteins of unknown function (Kolker et al. 2004). This uncertainty in gene function is reflective of how genome sequencing and annotation have outpaced the functional characterization of open reading frames and highlights that much is still to be discovered about the biology of prokaryotic genes.

One commonly encountered difficulty in the functional annotation of genes is an absence of an observable phenotype following their disruption. Given that natural selection and genetic drift lead to genome streamlining (Lynch 2006; Kuo and Ochman 2009; Kuo et al. 2009), and pseudogenes are rapidly lost from the genome (Lerat and Ochman 2005), it is unlikely that all of the uncharacterized genes truly lack a function. While many factors could contribute to a gene's apparent lack of function (e.g., the gene is not expressed

Published online: 01 February 2015

in the tested environment), one potential source is genetic redundancy within the genome. Genetic redundancy refers to the phenomenon whereby two genes or pathways are able to functionally complement the loss of the other (Zhang 2012). Genetic redundancy at both the pathway and single gene level is prevalent within eukaryotic genomes and has been extensively studied, particularly in yeast [see for example, *Saccharomyces cerevisiae* (Tong et al. 2001; Gu et al. 2003; Costanzo et al. 2010; Li et al. 2010), *Caenorhabditis elegans* (Tischler et al. 2006), and *Homo sapiens* (Hsiao and Vitkup 2008)]. Such redundancy can mask the fitness cost associated with the loss of a particular gene (Gu et al. 2003) and may make it difficult to identify the genetic determinates involved in a particular pathway through loss-of-function studies (Cutler and McCourt 2005).

Considering the streamlined nature of prokaryotic genomes, it may seem counter-intuitive to expect functionally redundant genes to be prevalent within a bacterial genome. Yet, there have been reports of functional redundancy within a broad range of microbial species and functions (for example, see Rabin and Stewart 1992; Belitsky et al. 2001; Elliot et al. 2003; Cheng et al. 2007; Xiao and Wall 2014). While these reports suggest that genetic redundancy through enzyme functional overlap is present within prokaryotic organisms, non-targeted large-scale studies are necessary to measure the extent of its prevalence. To date, large-scale studies have predominately focused on redundancy of metabolic pathways (Nakahigashi et al. 2009), predominately through *in silico* analyses (Ghim et al. 2005; Suthers et al. 2009; Wang and Zhang 2009). Recently, a synthetic genetic array was developed for *E. coli* (Butland et al. 2008), and genetic interaction maps have begun to be analyzed (Babu et al. 2014). Despite providing significant insight into the organism, neither of these approaches provides a direct examination of the presence of functionally redundant genes. In perhaps the largest examination of prokaryotic functionally redundant genes, Thomaides et al. (2007) experimentally examined 120 pairs of potentially redundant *Bacillus subtilis* genes identified through a bioinformatics analysis. These researchers found six pairs of redundant essential genes and a pair of redundant biosynthetic genes. However, the targeted nature of this study precludes reaching conclusions about the genome prevalence of genetic redundancy.

Sinorhizobium meliloti is a soil-dwelling, Gram-negative α-proteobacterium that enters into N_2-fixing endosymbiosis with several legumes belonging to the genera *Medicago*, *Melilotus*, and *Trigonella*. The genome of *S. meliloti* is multipartite, with all natural isolates containing at least an ~3.7 Mb chromosome, an ~1.4 Mb megaplasmid, and an ~1.7 Mb chromid (Galibert et al. 2001; Guo et al. 2009; Epstein et al. 2012). The identification of only two essential genes located outside of the chromosome (diCenzo et al. 2013; Milunovic et al. 2014) and their subsequent integration into the chromosome (diCenzo et al. 2013) facilitated the recent construction of an *S. meliloti* strain lacking 45 % of its genome through the removal of both pSymA and pSymB (diCenzo et al. 2014). The presence of this significantly reduced genome, accomplished through the removal of two out of three unlinked replicons, provided an opportunity to examine the level of genetic redundancy within the *S. meliloti* genome using a large-scale, non-targeted approach.

Here, we report the identification of chromosomal *S. meliloti* insertion mutants whose phenotype was dependent on the presence/absence of pSymA and pSymB (~45 % of the genome). Greater than 10 % of all mutations that abolished growth on minimal media were located within redundant genes. Redundancy was found in pathways for central carbon metabolism, transport, and amino acid biosynthesis. Characterization of these redundant gene pairs provided the first experimental confirmation of the function of several genes, adding to our understanding of *S. meliloti* metabolism.

Materials and methods

Media, growth conditions, and bacterial strains

All media [LB, LBmc (LB plus 2.5 mM $MgSO_4$ and 2.5 mM $CaCl_2$), TY, M9] were prepared as previously described (diCenzo et al. 2014). For growth of *S. meliloti*, all complex media (LB, LBmc, TY) were supplemented with 2 μM cobalt chloride (Cheng et al. 2011) and the M9 minimal medium with 5 μM thiamine–HCl (Finan et al. 1986; diCenzo et al. 2014). Antibiotic concentrations [streptomycin (Sm), spectinomycin (Sp), neomycin (Nm), kanamycin (Km), tetracycline (Tc), gentamicin (Gm), and chloramphenicol (Cm)] and growth conditions were as described elsewhere (diCenzo et al. 2014). Bacterial strains and plasmids are listed in Supplementary Table S1. When used as the sole carbon source, sucrose was added to a concentration of 10 mM, while glucose and succinate were added to 15 mM. When both were present, glucose and succinate were added to 10 mM each. Where stated, L-proline, L-arginine, and L-ornithine were added at a concentration of 1 mM.

Genetic manipulations

General DNA manipulations and recombinant techniques, bacterial matings, isolation of genomic *S. meliloti* DNA, and ΦM12 transductions were performed as previously described (Finan et al. 1984; Sambrook et al. 1989; Cowie et al. 2006; Milunovic et al. 2014). Where necessary,

plasmids were mobilized with the helper strain *E. coli* MT616 (pRK600) (Finan et al. 1986). Unless stated otherwise, cloning of DNA fragments into plasmids was performed with sequence- and ligation-independent cloning (SLIC) (Jeong et al. 2012). Oligonucleotides were ordered from Integrated DNA Technologies (IDT), and sequencing was performed by the MOBIX facility at McMaster University, Hamilton, Ontario, Canada. Oligonucleotide sequences used in this study are listed in Supplementary Table S2.

Transposon mutagenesis

The Tn*5*-B20 transposon (Simon et al. 1989) was transferred to *S. meliloti* ΔpSymAB (RmP2917) via the self-transmissible suicide vector pRK600::Tn*5*-B20. Transposon insertion mutants were selected on TY Sm Nm or LBmc Sm Nm and 40 independent matings were performed. Transposon insertion mutants were screened for an inability to grow on M9-sucrose and M9-cellobiose. Following the isolation of genomic DNA from mutants of interest, the location of the insertion was determined by sequencing out from the 3′ end of the transposon using the primer ML-1160 and mapping reads to the *S. meliloti* 1021 genome sequence (Galibert et al. 2001) available online at https://iant.toulouse.inra.fr/bacteria/annotation/cgi/rhime.cgi.

Complementation

A pLAFR1 cosmid library of *S. meliloti* 1021 genomic DNA (Friedman et al. 1982) was mated *en masse* into *S. meliloti* ΔpSymAB and complemented cells were selected for on M9-sucrose Sp Tc. Transconjugants were purified and complementing clones were conjugated into *E. coli* via a triparental mating with *E. coli* MT616 and rifampicin-resistant *E. coli* DH5α. Transconjugants were selected on either LB Rif (20 μg/uL) Tc or LB Tc and recovered on LB Tc at a frequency of ~10^{-8}/donor and ~10^{-8}/recipient. Cosmids were re-introduced into the appropriate *S. meliloti* ΔpSymAB transposon insertion mutant and screened for growth on M9-sucrose to confirm the complementation. The boarder ends of the *S. meliloti* DNA insert in the complementing pLAFR1 cosmids were identified through Sanger sequencing with primers P128 and P129.

Growth curves

Cultures grown in LBmc were washed once using carbon-free M9 medium, then resuspended and diluted into the M9 medium in which the growth curve was to be performed. All strains were grown in triplicate in 96-well microtitre plates. Growth conditions and analysis were as previously described (diCenzo et al. 2014).

Construction of pTH2919, a Tc^R *sacB* vector

To construct a Tc^R *sacB* vector, the Gm^R gene was removed from pJQ200mp18 (Quandt and Hynes 1993) through digestion with *Eco*RV and *Apa*LI. The Tc^R gene from pBBR1mcs-3 (Kovach et al. 1995) was PCR amplified with the primers DF001 and DF002, and introduced into the *Eco*RV-/*Apa*LI-digested pJQ200mp18, creating pTH2919.

Construction of plasmids for allelic replacement

Plasmids pTH2982 and pTH2983 were constructed to replace *sma0235* and *smb20003*, respectively, with the gentamicin resistance gene $aacC_4$. ~500 nucleotide fragments upstream and downstream of *sma0235* (primers: DF003/DF004 and DF005/DF006) and *smb20003* (primers: DF007/DF008 and DF009/DF010) were PCR amplified, and each pair was simultaneously cloned into *Xba*I-digested pTH2919, resulting in plasmids pTH2978 and pTH2979, respectively. These plasmids were digested with *Swa*I and ligated with a DNA fragment encoding $aacC_4$, which was PCR amplified from pHP45Ωaac (Blondelet-Rouault et al. 1997) using the primer DF011, resulting in the plasmids pTH2982 and pTH2983, respectively.

Construction of double deletions

To delete pSymA or pSymB genes that were putatively redundant with a chromosomal gene, Gm^R double recombinants of pTH2982 and pTH2983 into *S. meliloti* Rm2011 and the appropriate Rm2011 Tn*5*-B20 insertion mutants (pTH2982 into *edd*::Tn*5*-B20, pTH2983 into *proC*::Tn*5*-B20) were isolated. Plasmids were transferred into the appropriate recipient through a triparental mating on LB, and single recombinants were selected for on TY Sm Gm. Double recombinants were identified by screening for Tc sensitivity and for growth on LB Gm and on LB + 8 % sucrose.

Construction of pTH2987

pTH2987 was constructed to express *argH2 in trans* in *S. meliloti*. The coding region of *argH2* was PCR amplified from *S. meliloti* Rm2011 genomic DNA with primers DF012 and DF013, and was inserted into *Pac*I-digested pTH1931 (diCenzo et al. 2013), where *argH2* expression was driven from the P_{trc} promoter.

Sequence analysis

Nucleotide identity and amino acid similarity were determined following a global alignment of two sequences using the LALIGN algorithm provided on the ExPASy web server (Huang and Miller 1991; Gasteiger et al. 2003).

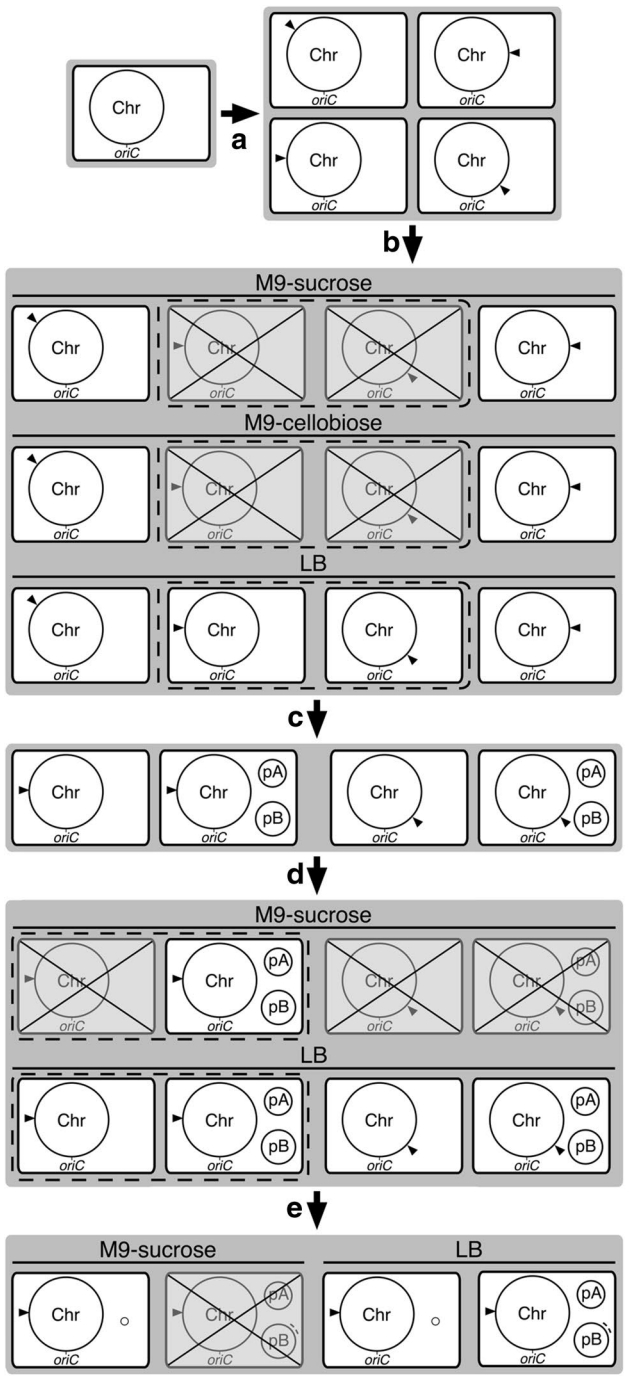

Fig. 1 Schematic representation of the experimental workflow. **a** ▸ A Tn*5*-B20 random insertion mutant library (represented by the *arrowheads*) of *S. meliloti* ΔpSymAB (RmP2917) was constructed. **b** The mutant library was replica plated (patched) on M9-sucrose, M9-cellobiose, and LB. Mutants unable to grow on M9-sucrose and M9-cellobiose were identified, as indicated by the *dashed boxes*. **c** The Tn*5*-B20 insertions from the mutants unable to grow on both minimal media were recombined (transduced) into naïve *S. meliloti* ΔpSymAB and wild-type *S. meliloti* Rm2011. **d** These recombinants were screened for the ability to grow on M9-sucrose. Transposon insertions resulting in no growth on M9-sucrose only in the absence of pSymA and pSymB were identified, as indicated by the *dashed boxes*. The location of these transposon insertions was determined via Sanger sequencing, as they were putatively located in genes with a redundant copy on pSymA/pSymB. **e** The loci on pSymA/pSymB that complemented the phenotypes of the transposon insertions were identified using two methods. *S. meliloti* ΔpSymAB transposon mutants were complemented with a genomic DNA cosmid library, as represented by the *small circle*. Additionally, *S. meliloti* Rm2011 transposon mutants were combined with large-scale deletions on pSymA/pSymB, as represented by the *dashed arch* adjacent to pSymB. **a–e** *Chr* chromosome, *pA* pSymA, *pB* pSymB

Results

Isolation of transposon insertions within redundant loci

To uncover redundancy between functions encoded on the pSymA or pSymB replicons and the chromosome, we sought to identify mutations whose phenotype was dependent on genome content. The experimental workflow is summarized in Fig. 1. ~9,300 Mutants from multiple Tn*5*-B20 metageneses of *S. meliloti* ΔpSymAB (ΔRmP2917) were screened for growth on M9-sucrose and M9-cellobiose, and 73 (~0.8 %) failed to grow on both media. Recombination via transduction of all 73 Tn*5*-B20 insertions back into *S. meliloti* ΔpSymAB showed linkage of the insertion with the inability to grow on M9-sucrose. At the same time, all 73 Tn*5*-B20 insertion mutations were recombined into the wild-type *S. meliloti* Rm2011 and screened for growth on M9-sucrose. Recombinants from 12 of the 73 crosses with *S. meliloti* Rm2011 grew on M9-sucrose. Recombination of these 12 Nm^R insertions back into *S. meliloti* Rm2011 and ΔpSymAB confirmed that these 12 insertions prevented growth on M9-sucrose only in the absence of pSymA and pSymB (Fig. 2). The 12 transposon insertions were found to be located within seven unique genes (*edd*, *proC*, *argH1*, *aglE*, *glgB1*, *pgk*, *argD*) via DNA sequencing (Fig. 3). Below, we summarize our characterization of the putatively redundant loci, and in many cases the identification of the redundant pSymA/pSymB loci. For the latter, we utilized a cosmid library carrying DNA from wild-type *S. meliloti* (Friedman et al. 1982) as well as a library of strains in which defined regions from pSymA or pSymB were deleted (Milunovic et al. 2014).

6-phosphogluconate dehydratase (*edd*)

Cosmids carrying a 21 kb region from nucleotide 125,586 to 147,721 of pSymA (note, all nucleotide positions for cosmid inserts and deletions are given relative to the Rm1021 reference genome) were found to complement the *edd-1*::Tn*5*-B20 allele. Confirming that the locus redundant with *edd* was located within this region, the introduction of the pSymA deletion ΔA105 (nt: 125,128–184,519) into RmP3099 (Rm2011, *edd-1*::Tn*5*-B20) resulted in

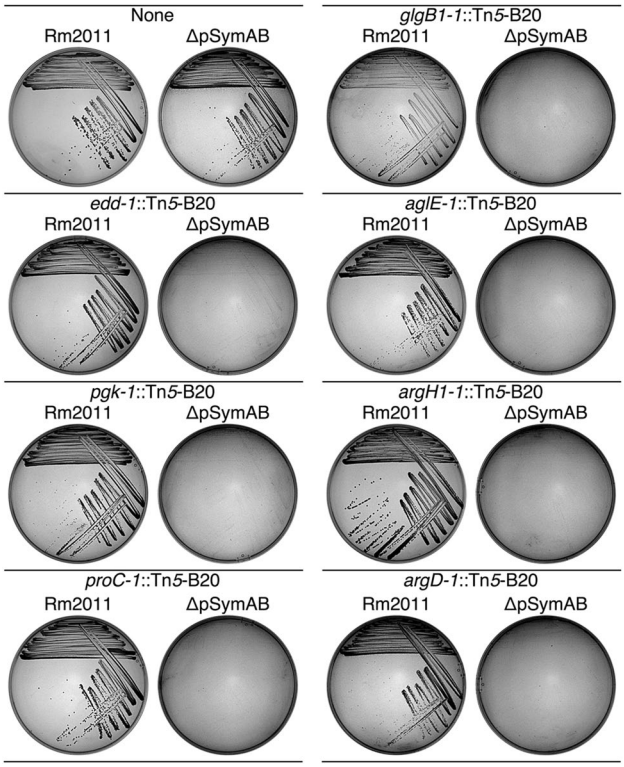

Fig. 2 Twelve transposon insertions, located within seven unique genes, prevented growth of *S. meliloti* on M9-sucrose agar plates in *S. meliloti* ΔpSymAB (RmP2917), but not Rm2011 (wild type). Pictures of the growth of representative alleles are shown. *S. meliloti* Rm2011 and the corresponding Tn*5*-B20 insertion mutants were incubated for 4 days, while *S. meliloti* ΔpSymAB and the corresponding Tn*5*-B20 insertion mutants were incubated for 5 days

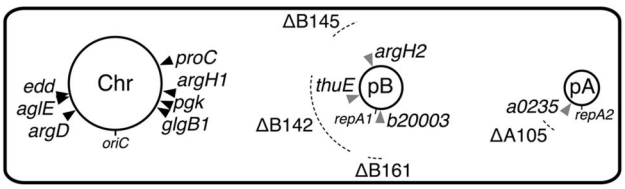

Fig. 3 The genomic location of the genes and deletions of interest in this study. *Black arrowheads* indicate the locations of genes disrupted by a Tn5-B20 transposon. *Gray arrowheads* signify the position of pSymA of pSymB genes able to complement the disruption of chromosomal genes. The *dotted curved lines* indicate the regions of pSymA and pSymB removed in the deletions used in this study. The position of the *oriC* of the chromosome, *repA1* on pSymB, and *repA2* on pSymA are pointed out for reference. *Chr* chromosome, *pA* pSymA, *pB* pSymB

extremely poor growth on M9-sucrose (data not shown). The most likely candidate gene to be redundant with *edd* within the 21 kb region of the complementing cosmids was *sma0235* [51.3 % nt identity to *edd*, 51.5 % aa (amino acid) similarity], annotated as a dihydroxy-acid dehydratase and 6-phosphogluconate dehydratase by InterPro (Hunter et al. 2012). Indeed, deletion of *sma0235* in RmP3099 resulted in little growth on M9-sucrose, while growth was indistinguishable from wild type on M9-succinate (Fig. 4a, b).

Pyrroline-5-carboxylate reductase (*proC*)

Complementation of *S. meliloti* RmP3104 (ΔpSymAB, *proC-1*::Tn*5*-B20) led to the isolation of clones carrying a 23 kb (nt: 1,669,658–9,262) region from pSymB. Introduction of the pSymB deletion ΔB181 (nt: 1,679,723–49,523) into RmP3103 (Rm2011, *proC-1*::Tn*5*-B20) resulted in no growth on M9-sucrose (data not shown), localizing the complementing locus to a 13 kb region. The *smb20003* gene (49.7 % nt identity to *proC*, 46.5 % aa similarity) within this region is annotated as a pyrroline-5-carboxylate reductase, as is *proC*. Deletion of *smb20003* in RmP3103 precluded growth on M9-sucrose unless supplemented with L-proline, L-ornithine, or L-arginine (Fig. 4c, d).

Argininosuccinate lyase (*argH1*)

There are two annotated copies of *argH* in the *S. meliloti* genome; *argH1* is on the chromosome and *argH2* (53.7 % nt identity to *argH1*, 68.5 % aa similarity) is on pSymB. Both were isolated on separate cosmids following complementation of *S. meliloti* RmP3110 (ΔpSymAB, *argH1-1*::Tn*5*-B20). No growth in M9-sucrose was observed when *argH2* was removed from RmP3109 (Rm2011, *argH1-1*::Tn*5*-B20) through the introduction of the pSymB deletion ΔB145 (nt: 635,940–744,320) (Fig. 4e). Growth on M9-sucrose was recovered following L-arginine supplementation (Fig. 4e). Furthermore, expression of the *argH2* gene *in trans* from pTH1931 complemented the arginine auxotrophy of RmP3242 (P3109, ΔB145) (Fig. 4e).

α-glucosides periplasmic substrate binding protein (*aglE*)

RmP3108 (ΔpSymAB, *aglE-1*::Tn*5*-B20) does not grow on sucrose (Fig. 2), but this strain grows well with cellobiose and glucose as carbon sources (data not shown). Complementation of the sucrose growth phenotype of *S. meliloti* RmP3108 identified library clones carrying a 26 kb region (nt: 327,602–354,008) from pSymB. Introduction of the pSymB deletion ΔB182 (nt: 122,108–466,499) into RmP3107 (Rm2011, *aglE*::Tn*5*-B20) confirmed the redundant locus to be within this region (data not shown). Located within this region is the *thu* ABC transporter, involved in uptake of several α-glucosides (Jensen et al. 2002). As it has previously been shown that the *thu* transporter is redundant with

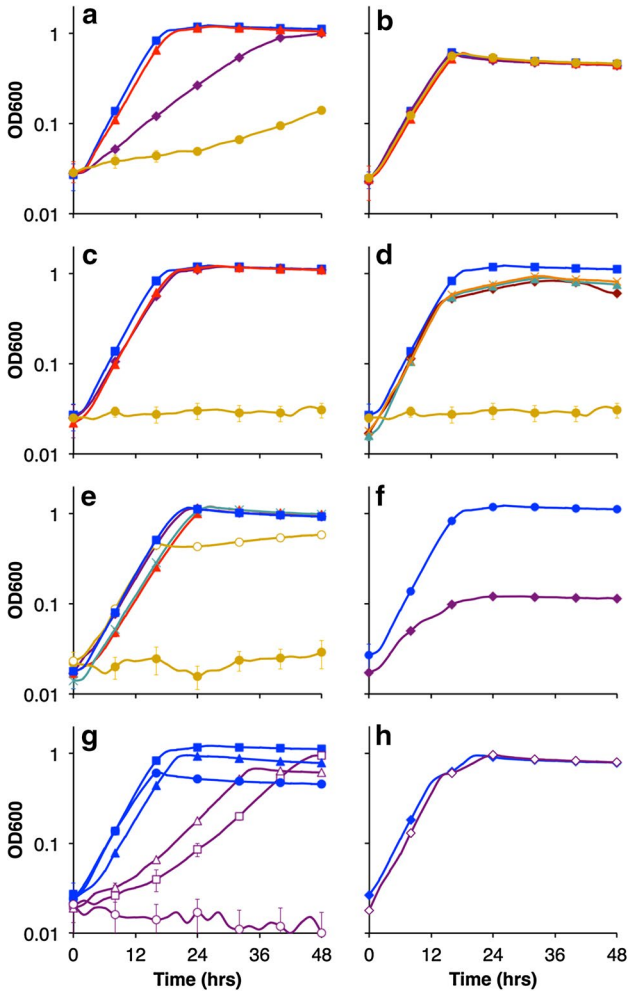

Fig. 4 Growth profiles of Tn*5*-B20 insertion mutants and associated strains. **a, b** Growth of wild type (*square*), *edd-1*::Tn*5*-B20 (*diamond*), Δ*sma0235*::*aacC*$_4$ (*triangle*), and *edd-1*::Tn*5*-B20 + Δ*sma0235*::*aacC*$_4$ (*circle*) in M9-sucrose (**a**) or M9-succinate (**b**). **c** Growth of wild type (*square*), *proC-1*::Tn*5*-B20 (*diamond*), Δ*smb20003*::*aacC*$_4$ (*triangle*), and *proC-1*::Tn*5*-B20 + Δ*smb20003*::*aacC*$_4$ (*circle*) in M9-sucrose. **d** Growth of wild type in M9-sucrose (*square*) and *proC-1*::Tn*5*-B20 + Δ*smb20003*::*aacC*$_4$ in M9-sucrose with no amino acid supplementation (*cirlce*), or L-proline (*diamond*), L-ornithine (*triangle*), or L-arginine (*cross*) supplementation. **e** Growth of wild type (*square*), *argH1-1*::Tn*5*-B20 (*diamond*), ΔB145 (*triangle*), *argH1-1*::Tn*5*-B20 + ΔB145 (*closed circle*), and *argH1-1*::Tn*5*-B20 + ΔB145 + *argH2 in trans* (*cross*) in M9-sucrose, as well as *argH1-1*::Tn*5*-B20 + ΔB145 in M9-sucrose supplemented with L-arginine (open circle). **f** Growth of wild type (*square*) and *glgB1-1*::Tn*5*-B20 (*diamond*) in M9-sucrose. **g** Growth of wild type (*solid symbols*) and *pgk-1*::Tn*5*-B20 (*open symbols*) in M9-sucrose (*square*), M9-glucose (*triangle*), and M9-succinate (*circle*). **h** Growth of wild type (*closed diamond*) and *pgk-1*::Tn*5*-B20 (*open diamond*) in M9-glucose + succinate. **a**–**h** Data *points* and *error bars* represent the average and standard deviation of triplicate samples. *Curves* were produced by plotting data *points* from readings every 2 h, with the data *points* shown for every 8 h

the *agl* ABC transporter for sucrose uptake (Willis and Walker 1999; Jensen et al. 2002), this redundancy was not further studied.

1,4-alpha-glucan branching enzyme (*glgB1*)

No significant growth of RmP3106 (ΔpSymAB, *glgB1-1*::Tn*5*-B20) was observed on M9-sucrose plates (Fig. 2). Slow growing colonies with a mucoid phenotype were observed on M9-sucrose plates for *S. meliloti* RmP3105 (Rm2011, *glgB1-1*::Tn*5*-B20) (Fig. 2), while little growth was observed for this strain in liquid M9-sucrose medium (Fig. 4f). The five complementing clones that were analyzed carried the wild-type *glgB1* and we therefore failed to identify a pSymA or pSymB locus redundant with *glgB1*. However, transduction of *glgB1-1*::Tn*5*-B20 into strains lacking either pSymA or pSymB indicated that the severity of the phenotype of the *glgB1-1*::Tn*5*-B20 mutation was increased by the lack of either pSymA or pSymB (Supplementary Fig. S1).

Phosphoglycerate kinase (*pgk*)

All five of the isolated cosmids that complemented *S. meliloti* RmP3102 (ΔpSymAB, *pgk-1*::Tn*5*-B20)-encoded *pgk*. Thus, no redundant locus was identified via complementation. Intriguingly, transduction of *pgk-1*::Tn*5*-B20 into strains lacking just pSymA or pSymB revealed the phenotype of the *pgk-1*::Tn*5*-B20 mutation was impacted by the absence of either pSymA or pSymB (Fig. 1, Supplementary Fig. S1). Growth studies revealed that while RmP3101 (Rm2011, *pgk-1*::Tn*5*-B20) did grow in M9-sucrose, it was significantly impaired (Fig. 4g). Slow growth of RmP3101 relative to the wild type was also observed in M9-glucose, while no growth was observed in M9-succinate (Fig. 4g). However, growth was largely indistinguishable from that of the wild type when grown in M9 with both glucose and succinate (Fig. 4h).

Acetylornithine aminotransferase (*argD*)

A Tn*5*-B20 insertion within *argD* was isolated that resulted in L-arginine auxotrophy in the ΔpSymAB strain, but not the wild-type *S. meliloti* Rm2011 (Fig. 2 and data not shown). Transduction of *argD-1*::Tn*5*-B20 into strains lacking either pSymA or pSymB revealed the arginine auxotrophy of *argD-1*::Tn*5*-B20 is only observed in the absence of pSymB (Supplementary Fig. S1). However, we did not examine the precise location of the redundant locus.

Discussion

Identification and characterization of functionally redundant genes

10–15 % of transposon insertions in the *S. meliloti* chromosome generated phenotypes on M9-sucrose that were dependent on the absence of the pSymA or pSymB replicons. The identified chromosomal genes with a functionally redundant gene on pSymA or pSymB were involved in amino acid biosynthesis (*proC*, *argH1*, and *argD*), central carbon metabolism (*edd*), and transport (*aglE*).

S. meliloti does not possess a complete Embden–Meyerhof–Parnas (EMP) glycolytic pathway, as it does not encode a phosphofructokinase enzyme (Arias et al. 1979; Galibert et al. 2001). Instead, glycolytic substrates such as sucrose and glucose are metabolized via the Entner–Doudoroff (ED) pathway to pyruvate and glyceraldehyde-3-phosphate, which is further metabolized to form a second pyruvate via the lower half of the EMP pathway (Fig. 5; Stowers 1985; Geddes and Oresnik 2014). Gluconeogenesis proceeds via the EMP pathway in *S. meliloti* and does not involve the ED pathway (Fig. 5; Finan et al. 1988). The gene *edd* encodes a 6-phosphogluconate dehydratase, which catalyzes the first step unique to the ED pathway. Despite previous study on carbon metabolism in *S. meliloti* (e.g., Finan et al. 1988), mutants defective in 6-phosphogluconate dehydratase activity have never been isolated. Consistent with this, disrupting *edd* decreases, but does not prevent, growth on sucrose (Fig. 4a). However, combining this disruption with a deletion of a pSymA-encoded gene, *sma0235*, results in a synthetic phenotype and very little growth with sucrose as the sole carbon source (Fig. 4a). As *sma0235* shows homology to dehydratases, in particular, dihydroxy-acid and 6-phosphogluconate dehydratases, it is likely that the gene products of *edd* and *sma0235* share overlapping enzymatic specificity. However, while Sma0235 seemingly has activity with 6-phosphogluconate, this is unlikely to be the primary substrate of this enzyme for two reasons: one, *sma0235* in unable to completely complement the disruption of *edd*; two, *sma0235* is the third gene of a three-gene operon that also consists of an epimerase (the first gene, *sma0241*) and a dehydrogenase (the second gene, *sma0237*), suggesting that this operon is involved in a metabolic pathway distinct from that of the ED pathway.

L-proline synthesis from L-glutamate occurs via a three-enzyme pathway, with the final step the conversion of Δ^1-pyrroline-5-carboxylate to L-proline catalyzed by ProC (Fig. 6). However, disrupting the chromosomal *proC* gene had no phenotype unless combined with a deletion of the pSymB-encoded gene *smb20003* (Fig. 4c). As *smb20003* shares the same annotation of

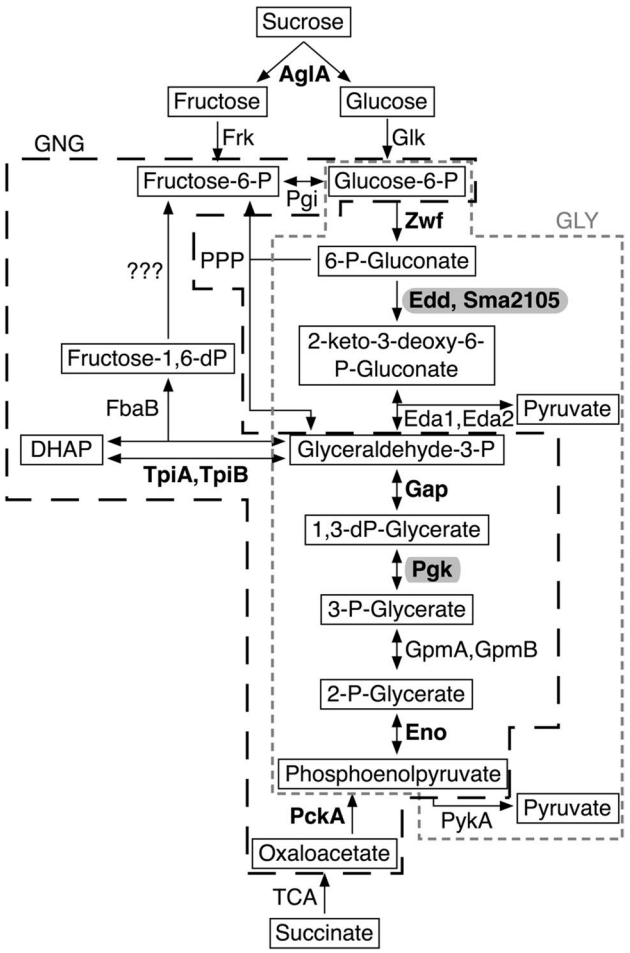

Fig. 5 Schematic of the glycolytic (GLY) and gluconeogenic (GNG) pathways of *S. meliloti*. The *short dashed gray box* indicates the steps of the GLY pathway, while the *long dashed black box* indicates the steps of the GNG pathway. Proteins catalyzing each step are indicated. Proteins with experimental evidence supporting their designated molecular function are indicated with *boldface*, while the rest are classified based solely on annotation. Proteins of interest in this study are highlighted with a *gray box*. *TCA* tricarboxylic acid cycle, *PPP* pentose phosphate pathway, *DHAP* dihydroxyacetone phosphate, *???* enzyme unknown. For a review of carbon metabolism in *S. meliloti*, please refer to Geddes and Oresnik (2014)

proC, a Δ^1-pyrroline-5-carboxylate reductase, it is likely that the synthetic effect observed in the double mutant is due to the elimination of enzymatic, not pathway, redundancy. Transcriptional start site mapping indicates that *smb20003* is transcribed as a monocistronic mRNA (Schlüter et al. 2013), while a reciprocal best-hit strategy suggests that both ProC and Smb20003 orthologs are present throughout the *Rhizobiaceae* family (data not shown). In addition to redundancy of *proC*, two genes are annotated as encoding the ProB enzyme on the *S. meliloti* chromosome, while *proA* appears to present in a single copy (Fig. 6). This is similar to the situation in *B. subtilis*, which encodes two orthologs of *proB*, three of *proC*, and

only one ortholog of *proA* (Belitsky et al. 2001). While the two *proB* and three *proC* homologs in *B. subtilis* can complement each other, the transcriptional up-regulation of each homolog occurs in response to unique stimuli (Belitsky et al. 2001).

In addition to synthesis of L-proline from L-glutamate, alternate L-proline biosynthetic pathways exist in prokaryotic species, such as the conversion of L-ornithine to L-proline through the activity of ornithine cyclodeaminases (Fig. 6). Previous work has identified ornithine cyclodeaminase activity in *S. meliloti* cell extracts (Soto et al. 1994), and four genes (*ocd*, *eutC*, *sma0486*, *sma1871*) are annotated as encoding ornithine cyclodeaminases in *S. meliloti* Rm2011. Indeed, auxotrophy of the *proC, smb20003* double mutant was eliminated by adding exogenous L-ornithine (Fig. 4d). Furthermore, exogenous L-arginine also eliminated the proline auxotrophy (Fig. 4d), presumably as L-arginine can be converted to L-ornithine either through an arginase (i.e., ArgI1 or ArgI2) or the ArcABC catabolic pathway (Fig. 6). Similarly, L-arginine and L-ornithine can be converted to L-proline in *B. subtilis*; however, unlike in *S. meliloti*, this pathway proceeds through a ProC-catalyzed final step (Belitsky et al. 2001). While we are unsure why such extensive enzymatic and pathway level redundancy in L-proline biosynthesis exists, its presence in highly diverse species suggests it may provide a fitness advantage to the cell.

Two of the redundant gene pairs encode proteins involved in the eight-step pathway converting L-glutamate to L-arginine (Fig. 6). In the final step of this pathway, L-argininosuccinate lyase (ArgH) converts L-argininosuccinate to L-arginine. Two *argH* genes are annotated in the *S. meliloti* genome, one on the chromosome (*argH1*) and one on pSymB (*argH2*). Consistent with this function being redundant, disruption/deletion of both genes renders *S. meliloti* an arginine auxotroph, whereas the presence of either gene individually is sufficient for arginine prototrophy (Fig. 4e). The gene *argH2* is the first gene of an operon encoding an ATP transporter (*smb21095–smb21097*) and two hypothetical genes (*smb21098*, *smb21100*). Previous work has shown that transcription of this ATP transporter is induced in the presence of L-citrulline (Mauchline et al. 2006), whose conversion to L-arginine or catabolism through the urea cycle involves an L-argininosuccinate lyase (ArgH)-catalyzed step (Kanehisa et al. 2014). Thus, the presence of two copies of *argH* in the *S. meliloti* genome presumably allows differential regulation and synthesis of the L-argininosuccinate lyase isozymes: induction of *argH1* by a lack of L-arginine and the upregulation of *argH2* by the presence of L-citrulline.

The second Tn*5*-B20 insertion within a redundant arginine biosynthetic locus was situated near the 3′ end of *argD*, encoding an L-acetylornithine aminotransferase. The

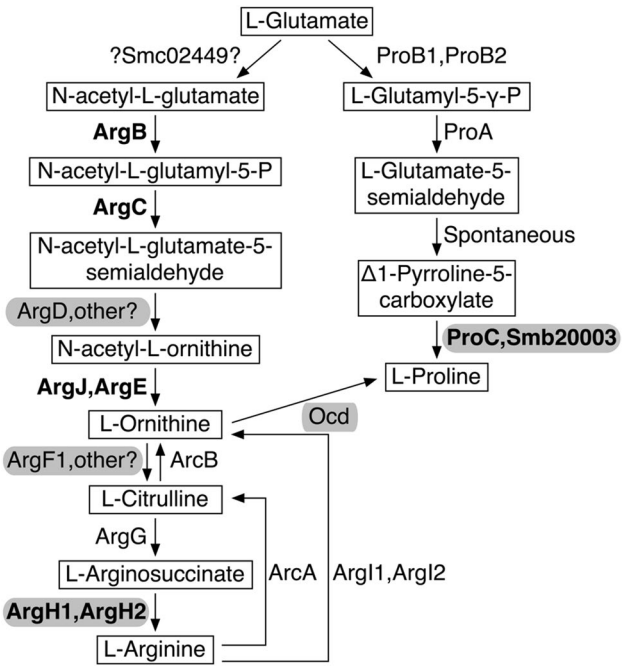

Fig. 6 Schematic of the L-arginine and L-proline biosynthetic pathways of *S. meliloti*. Proteins catalyzing each step are indicated. Proteins with experimental evidence supporting their designated molecular function are indicated with *boldface*, while the rest are classified based solely on annotation. The enzyme catalyzing the first reaction in the L-arginine pathway is unknown; however, it may be Smc02449 based on genome context and homology to the *Corynebacterium glutamicum* NAG synthase (data not shown). Proteins of interest in this study are highlighted with a *gray box*. For a review of amino acid metabolism in rhizobia, please refer to Dunn (2014)

gene *argD* is situated less than 100 base pairs upstream of *argF1*, annotated as encoding an L-ornithine carbamoyltransferase that is presumably also involved in the arginine biosynthetic pathway (Fig. 6). Two independent RNA-seq experiments suggest the presence of transcriptional start sites upstream of *argD* and upstream of *argF1* (Schlüter et al. 2013; Milunovic et al. 2014); however, a promoter motif was only identified upstream of *argD* (Schlüter et al. 2013). Thus, it is unclear whether *argD* and *argF1* form a bicistronic operon. Furthermore, the Tn*5*-B20 transposon inserted downstream of all codons encoding conserved residues in ArgD, meaning a functional version of ArgD, may still have been translated. The redundant locus (or loci) is located on pSymB (Supplementary Fig. S1); however, we did not attempt to identify the precise gene(s), and therefore cannot state whether the *argD-1*::Tn*5*-B20 phenotype is associated with the loss of ArgD, ArgF1, or both. Nevertheless, we note that ArgD shows similarity to five pSymB aminotransferases (as well as five encoded by the chromosome and four encoded by pSymA), while ArgF1 shows similarity to one pSymB-encoded protein (as well as one pSymA-encoded protein).

Additional genome-dependant phenotypes

Phosphoglycerate kinase (Pgk) catalyzes the reversible conversion of 1,3-diphosphate-glycerate and 3-phosphoglycerate, and is thus essential for growth of *S. meliloti* with gluconeogenic substrate (Fig. 5; Finan et al. 1988). While Pgk is also involved in the glycolytic pathway, it is not absolutely required as one of the pyruvic acid products of the ED pathway bypasses Pgk (Fig. 5; Finan et al. 1988). Here, it was observed that *pgk-1*::Tn*5*-B20 mutants grew slowly with glycolytic substrate (sucrose and glucose) as sole carbon sources only when pSymA and pSymB were a part of the genome (Figs. 2, 4g, Supplementary Fig.S1). We hypothesize that disruption of *pgk* leads to a buildup of glycolytic intermediates that, unless removed through pSymA- and pSymB-dependent pathways, results in an eventual cessation of glycolysis and no formation of either pyruvate (Fig. 5). Similarly, it was previously suggested that pathways removing glycolytic intermediates explain, in part, why *pgk* mutations in *Pseudomonas aeruginosa* do not prevent growth on glycolytic substrate, unlike in *E. coli* (Banergee et al. 1987).

S. meliloti predominately stores carbon as two polymers, poly-3-hydroxybutyrate and glycogen, which appear to have separate functions during N_2-fixing symbiosis with legumes (Wang et al. 2007). Glycogen is synthesized via a three-enzyme pathway consisting of a glucose-1-phosphate adenylyltransferase (GlgC), glycogen synthetase (GlgA), and a glycogen branching enzyme (GlgB) (Kanehisa et al. 2014). Little growth of *glgB1-1*::Tn*5*-B20 mutants was observed in liquid M9-sucrose medium (Fig. 4f), although slow growth was observed on M9-sucrose plates unless pSymA and pSymB were removed (Fig. 2, Supplementary Fig. S1). No redundant pSymA- or pSymB-encoded locus was identified through complementation with the genomic DNA cosmid library. Additionally, combining the *glgB1-1*::Tn*5*-B20 allele with pSymB deletions of candidate genes (*glgB2*, *bdhA*, *bhbA*, and the exopolysaccharide biosynthetic cluster) failed to reveal synthetic interactions (data not shown). Hence, the mechanistic basis for the pSymA or pSymB redundancy remains unknown.

Genetic redundancy in bacterial organisms

Redundant gene pairs encoded 10–15 % of functions required for growth on minimal M9-sucrose medium. Considering the potential for intra-chromosomal redundancy, which would not have been detected in the present study, we posit that this is an underestimate of the total genetic redundancy encoding processes essential for growth on M9-sucrose. These results illustrate that genetic redundancy may be extensive within large bacterial genomes, an observation that raises several theoretical questions and has significant practical implications.

It may not be particularly surprising to find significant genetic redundancy in the *S. meliloti* genome given that ~40 % of the predicted genes belong to gene families, although few of these are the result of recent duplications (Galibert et al. 2001). While it is common for large bacterial genomes to have a high percentage of genes belonging to gene families (Glass et al. 2005), our data show that one must be careful to extrapolate from amino acid similarity to genetic redundancy and functional complementation. As an example, Edd shows significant amino acid similarity to six proteins (IlvD4, IlvD2, IlvD3, IlvD1, AraF, Sma0235). Yet, only Sma0235 is able to partially complement *edd* mutants despite being the most dissimilar protein to Edd of the six (Edd: Sma0235, 51.5 % aa similarity; Edd: others, 54.4–54.6 % aa similarity). In many cases, two or more proteins may have similar molecular functions, but serve unique biological roles and thus do not complement mutations of the others, such as the five *S. meliloti* copper ABC transporters (Patel et al. 2014). Thus, while bioinformatics provides insight into the potential level of genetic redundancy in a genome, an experimental approach is required for a full understanding.

Considering the streamlined nature of prokaryotic genomes, a high level of genetic redundancy seems counter-intuitive and we wonder why such redundancy exists. Genetic redundancy could simply be the result of a recent gene duplication and an insufficient amount of time for purifying selection to lead to the loss of one copy. However, this is unlikely to be true in examples where the redundant genes are conserved (e.g., *proC*). Redundant genes can serve to increase the rate of production of highly expressed gene products, such as is posited for rRNA (Bremer 1975; Condon et al. 1995) and experimentally increasing gene copy number can increase gene expression (Janczarek et al. 2009). Functionally redundant proteins could also be maintained if they show partial substrate overlap. For example, both AglE and ThuE bind sucrose, but only AglE binds cellobiose while ThuE uniquely binds palatinose (Jensen et al. 2002; Ampomah et al. 2013). Moreover, differential regulation could serve as a driving force for the maintenance of redundant genes by optimizing metabolic versatility, similar to how the redundant proline biosynthetic genes in *B. subtilis* are up-regulated in response to different stimuli (Belitsky et al. 2001).

Practical implications/consequences of genetic redundancy

Genetic redundancy in prokaryotic genomes has practical implications/consequences, primarily the masking of phenotypes during loss-of-function studies. This would limit our ability to identify the genetic determinants of a pathway or to ascertain the biological role of a gene through a mutagenesis-based approach. For example, the last step of

gluconeogenesis in *S. meliloti* (Fig. 5) remains uncharacterized despite intensive characterization of this pathway (e.g., Finan et al. 1988), perhaps reflective of redundancy in this step. Phenotype masking through genetic redundancy also impacts the conclusions drawn from high-throughput transposon mutagenesis studies. The use of saturation transposon mutagenesis has been used in the global mapping of essential genes and identification of the minimal bacterial genome (e.g., Jacobs et al. 2003; Glass et al. 2005; Christen et al. 2011), bacterial virulence (e.g., Wu et al. 2006; Gawronski et al. 2009; van Opijnen and Camilli 2012), and other biological processes (e.g., Stewart et al. 2004; Cameron et al. 2008). While such studies provide invaluable genetic information into many biological processes, it must be recognized that greater than 10 % of target genes may be missed due to genetic redundancy. A failure to consider redundantly encoded functions could lead to an incomplete understanding of many important biological processes. Not only is this an issue in the sense of fully understanding the biology of the cell, an incomplete understanding of biological pathways will hinder our ability to use synthetic biology to construct designer cell factories.

Acknowledgments Special thanks to Arlene Sutherland and the students in the Molecular Biology 3V03 course at McMaster University, who isolated many of the mutants used in this study. This work was supported by the Natural Sciences and Engineering Research Council of Canada through grants to T. M. F and an NSERC CGS scholarship to G. C. D.

References

Ampomah OY, Avetisyan A, Hansen E, Svenson J, Juser T, Jensen JB, Bhuvaneswari TV (2013) The *thuEFGKAB* operon of rhizobia and *Agrobacterium tumefaciens* codes for transport of trehalose, maltitol, and isomers of sucrose and their assimilation through the formation of their 3-keto derivatives. J Bacteriol 195:3797–3807

Arias A, Cerveñansky C, Gardiol A, Martínez-Drets G (1979) Phosphoglucose isomerase mutant of *Rhizobium meliloti*. J Bacteriol 137:409–414

Babu M, Arnold R, Bundalovic-Torma C, Gargarinova A, Wong KS, Kumar A, Stewart G, Samanfar B, Aoki H, Wagih O, Vlasblom J, Phanse S, Lad K, Yeou Hsiung Yu A, Graham C, Jin K, Brown E, Gloshani A, Kim P, Moreno-Hagelsieb G, Greenblatt J, Joury WA, Parkinson J, Emili A (2014) Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in *Escherichia coli*. PLoS Genet 10:e1004120

Banergee PC, Darzins A, Maitra PK (1987) Gluconeogenic mutations in *Pseudomonas aeruginosa*: genetic linkage between fructose-bisphosphate aldolase and phosphoglycerate kinase. J Gen Microbiol 133:1099–1107

Belitsky BR, Brill J, Bremer E, Sonenshein AL (2001) Multiple genes for the last step of proline biosynthesis in *Bacillus subtilis*. J Bacteriol 183:4389–4392

Blondelet-Rouault MH, Weiser J, Lebrihi A, Branny P, Pernodet JL (1997) Antibiotic resistance gene cassettes derived from the omega interposon for use in *E. coli* and *Streptomyces*. Gene 190:315–317

Bremer H (1975) Parameters affecting the rate of synthesis of ribosomes and RNA polynerase in bacteria. J Theor Biol 53:115–124

Butland G, Babu M, Díaz-Mejía JJ, Bohdana F, Phanse S, Gold B, Yang W, Li J, Gagarinova AG, Pogoutse O, Mori H, Wanner BL, Lo H, Wasniewski J, Christopolous C, Ali M, Venn P, Safavi-Naini A, Sourour N, Caron S, Choi J-Y, Laigle L, Nazarians-Armavil A, Deshpande A, Joe S, Datsenko KA, Yamamoto N, Andrews BJ, Boone C, Ding H, Sheikh B, Moreno-Hagelseib G, Greenblatt JF, Emili A (2008) eSGA: *E. coli* synthetic genetic array analysis. Nat Methods 5:789–795

Cameron DE, Urbach JM, Mekalanos JJ (2008) A defined transposon mutant library and its use in identifying motility genes in *Vibrio cholerae*. Proc Natl Acad Sci USA 105:8736–8741

Cheng J, Sibley CD, Zaheer R, Finan TM (2007) A *Sinorhizobium meliloti minE* mutant has an altered morphology and exhibits defects in legume symbiosis. Microbiology 153:375–387

Cheng J, Poduska B, Morton RA, Finan TM (2011) An ABC-type cobalt transport system is essential for growth of *Sinorhizobium meliloti* at trace metal concentrations. J Bacteriol 193:4405–4416

Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, Fero MJ, McAdams HH, Shapiro L (2011) The essential genome of a bacterium. Mol Syst Biol 7:528

Condon C, Liveris D, Squires C, Schwartz I, Squires CL (1995) rRNA operon multiplicity in *Escherichia coli* and the physiological implications of *rrn* inactivation. J Bacteriol 177:4152–4156

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S, Prinz J, St. Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin Z-Y, Liang W, Marback M, Paw J, San Luis B-J, Shuteriqi E, Tong AHY, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras A-C, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C (2010) The genetic landscape of a cell. Science 327:425–431

Cowie A, Cheng J, Sibley CD, Fong Y, Zaheer R, Patten CL, Morton RM, Golding GB, Finan TM (2006) An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. Appl Environ Microbiol 72:7156–7167

Cutler S, McCourt P (2005) Dude, where's my phenotype? Dealing with redundancy in signaling networks. Plant Physiol 138:558–559

diCenzo G, Milunovic B, Cheng J, Finan TM (2013) The tRNAarg 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 195:202–212

diCenzo GC, MacLean AM, Milunovic B, Golding GB, Finan TM (2014) Examination of prokaryotic multipartite genome evolution through experimental genome reduction. PLoS Genet 10:e1004742

Dunn MF (2014) Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. Crit Rev Microbiol. doi:10.3109/1040841X.2013.856854

Elliot MA, Karoonuthaisiri N, Huang J, Bibb MJ, Cohen SN, Kao CM, Buttner MJ (2003) The chaplins: a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in *Streptomyces coelicolor*. Genes Dev 17:1727–1740

Epstein B, Branca A, Mudge J, Bharti AK, Briskine R, Farmer AD, Sugawara M, Young ND, Sadowsky MJ, Tiffin P (2012) Population genomics of the facultatively mutualistic bacteria *Sinorhizobium meliloti* and *S. medicae*. PLoS Genet 8:e1002868

Finan TM, Hartweig E, LeMieux K, Bergman K, Walker GC, Signer ER (1984) General transduction in *Rhizobium meliloti*. J Bacteriol 159:120–124

Finan TM, Kunkel B, De Vos GF, Signer ER (1986) Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. J Bacteriol 167:66–72

Finan TM, Oresnik IJ, Bottacin A (1988) Mutants of *Rhizobium meliloti* defective in succinate metabolism. J Bacteriol 170:3396–3403

Friedman AM, Long SR, Brown SE, Buikema WJ, Ausubel FM (1982) Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene 18:289–296

Galibert F, Finan TM, Long SR, Pühler A, Abola AP, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dréano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandex-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl T, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thébault P, Vendenbol M, Vorhölter F-J, Weidner S, Wells DH, Wong K, Yeh KC, Batut J (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293:668–672

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788

Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. Proc Natl Acad Sci USA 106:16422–16427

Geddes BA, Oresnik IJ (2014) Physiology, genetics, and biochemistry of carbon metabolism in the alphaproteobacterium *Sinorhizobium meliloti*. Can J Microbiol 60:491–507

Ghim CM, Goh KI, Kahng B (2005) Lethality and synthetic lethality in the genome-wide metabolic network of *Escherichia coli*. J Theor Biol 237:401–411

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA III, Smith HO, Venter JC (2005) Essential genes of a minimal bacterium. Proc Natl Acad Sci USA 103:425–430

Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RM, Li W-H (2003) Role of duplicate genes in genetic robustness against null mutations. Nature 421:63–66

Guo H, Donini P, Sun S, Eardly B, Finan T, Xu J (2009) Genome variation in the symbiotic nitrogen-fixing bacterium *Sinorhizobium meliloti*. Genome 52:862–875

Hsiao T-L, Vitkup D (2008) Role of duplicate genes in robustness against deleterious human mutations. PLoS Genet 4:e1000014

Huang X, Miller W (1991) A time-efficient, linear-space local similarity algorithm. Adv Appl Math 12:337–357

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duguenne L, Finn RD, Fraser M, Gough J, Haft D, Julo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, McMenamin C, Mi H, Mutowo-Muellenet P, Mulder N, Natale D, Orengo C, Pesseat S, Punta M, Quinn AF, Rivoire C, Sangrador-Vegas A, Selengut JD, Sigrist CJA, Scheremetjew M, Tate J, Thimmajanarthanan M, Thomas PD, Wu CH, Yeats C, Yong S-Y (2012) InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res 40:D306–D312

Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Jaugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R, Chun-Rong L, Guenthner D, Bovee D, Olson MV, Manoil C (2003) Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. Proc Natl Acad Sci USA 100:14339–14344

Janczarek M, Jaroszuk-Ściseł J, Skorupska A (2009) Multiple copies of *rosR* and *pssA* genes enhance exopolysaccharide production, symbiotic competitiveness and clover nodulation in *Rhizobium leguminosarum* bv. Trifolii. Antonie Van Leeuwenhoek 96:471–486

Jensen JB, Peters NK, Bhuvaneswari TV (2002) Redundancy in periplasmic binding protein-dependent transport systems for trehalose, sucrose, and maltose in *Sinorhizobium meliloti*. J Bacteriol 184:2978–2986

Jeong J-Y, Yim H-S, Ryu J-Y, Lee HS, Lee J-H, Seen D-S, Kang SG (2012) One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. Appl Environ Microbiol 78:5440–5443

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 42:D199–D205

Kolker E, Makarova KS, Shabalina S, Picone AF, Purvine S, Holzman T, Cherny T, Armbruster D, Munson RS Jr, Kolesov G, Frishman D, Galperin MY (2004) Identification and functional analysis of "hypothetical" genes expressed in *Haemophilus influenzae*. Nucleic Acids Res 32:2353–2361

Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM, Peterson KM (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene 166:175–176

Kuo C-H, Ochman H (2009) Deletional bias across the three domains of life. Genome Biol Evol 1:145–152

Kuo C-H, Moran NA, Ochman H (2009) The consequences of genetic drift for bacterial genome complexity. Genome Res 19:1450–1454

Lerat E, Ochman H (2005) Recognizing the pseudogenes in bacterial genomes. Nucleic Acids Res 33:3125–3132

Li J, Yuan Z, Zhang Z (2010) The cellular robustness by genetic redundancy in budding yeast. PLoS Genet 6:e1001187

Lynch M (2006) Streamlining and simplification of microbial genome architecture. Annu Rev Microbiol 60:327–349

Maglott D, Ostell J, Pruitt KD, Tatusova T (2011) Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 39:D52–D57

Mauchline TH, Fowler JE, East AK, Sartor AL, Zaheer R, Hosie AHF, Poole PS, Finan TM (2006) Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. Proc Natl Acad Sci USA 103:17933–17938

Milunovic B, diCenzo GC, Morton RA, Finan TM (2014) Cell growth inhibition upon deletion of four toxin-antitoxin loci from the megaplasmids of *Sinorhizobium meliloti*. J Bacteriol 196:811–824

Nakahigashi K, Toya Y, Ishii N, Soga T, Hasegawa M, Watanabe H, Takai Y, Honma M, Mori H, Tomita M (2009) Systematic phenome analysis of Escherichia coli multiple-knockout mutants reveals hidden reactions in central carbon metabolism. Mol Syst Biol 5:306

Patel SJ, Padilla-Benavides T, Collins JM, Argüello JM (2014) Functional diversity of five homologous Cu^{+}-ATPases present in *Sinorhizobium meliloti*. Microbiology 160:1237–1251

Quandt J, Hynes MF (1993) Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. Gene 127:15–21

Rabin RS, Stewart V (1992) Either of two functionally redundant sensor proteins, *NarX* and *NarQ*, is sufficient for nitrate regulation in *Escherichia coli* K-12. Proc Natl Acad Sci USA 89:8419–8423

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Schlüter J-P, Reinkensmeier J, Barnett MJ, Lang C, Krol E, Giegerich R, Long SR, Becker A (2013) Global mapping of transcription start sites and promoter motifs in the symbiotic α-proteobacterium *Sinorhizobium meliloti* 1021. BMC Genom 14:156

Simon R, Quandt J, Klipp W (1989) New derivatives of transposon Tn*5* suitable for mobilization of replicons, generation of operon fusions and induction of genes in Gram-negative bacteria. Gene 80:161–169

Soto MJ, van Dillewijn P, Olivares J, Toro N (1994) Ornithine cyclodeaminase activity in *Rhizobium meliloti*. FEMS Microbiol Lett 119:209–214

Stewart PE, Hoff J, Fischer E, Krum JG, Rosa RA (2004) Genome-wide transposon mutagenesis of *Borrelia burgdorferi* for identification of phenotypic mutants. Appl Environ Microbiol 70:5973–5979

Stowers MD (1985) Carbon metabolism in *Rhizobium* species. Annu Rev Microbiol 39:89–108

Suthers PF, Zomorrodi A, Maranas CD (2009) Genome-scale gene/reaction essentiality and synthetic lethality analysis. Mol Syst Biol 5:301

Thomaides HB, Davison EJ, Burston L, Johnson H, Brown DR, Hunt AT, Errington J, Czaplewski L (2007) Essential bacterial functions encoded by gene pairs. J Bacteriol 189:591–602

Tischler J, Lehner Ben, Chen N, Fraser AG (2006) Combinatorial RNA interference in *Caenorhabditis elegans* reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. Genome Biol 7:R69

Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CWV, Bussey H, Andrews B, Tyers M, Boone C (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294:2364–2368

van Opijnen T, Camilli A (2012) A fine scale phenotype-genotype virulence map of a bacterial pathogen. Genome Res 22:2541–2551

Wang Z, Zhang J (2009) Abundant indispensable redundancies in cellular metabolic networks. Genome Biol Evol 1:23–33

Wang C, Saldanha M, Sheng X, Shelswell KJ, Walsh KT, Sobral BWS, Charles TC (2007) Roles of poly-3-hydroxybutyrate (PHB) and glycogen in symbiosis of *Sinorhizobium meliloti* with *Medicago* sp. Microbiology 153:388–398

Willis LB, Walker GC (1999) A novel *Sinorhizobium meliloti* operon encodes an α-glucosidase and a periplasmic-binding-protein-dependent transport system for α-glucosides. J Bacteriol 181:4176–4184

Wu Q, Pei J, Turse C, Ficht TA (2006) Mariner mutagenesis of *Brucella melitensis* reveals genes with previously uncharacterized roles in virulence and survival. BMC Microbiol 6:102

Xiao Y, Wall D (2014) Genetic redundancy, proximity, and functionality of *lspA*, the target of antibiotic TA, in the *Myxococcus xanthus* producer strain. J Bacteriol 196:1174–1183

Zhang J (2012) Genetic redundancies and their evolutionary maintenance. In: Soyer OS (ed) Evolutionary systems biology. Springer, New York, pp 279–300