The Fitness Cost of Horizontal Gene Transfer

David A. Baltrus
School of Plant Sciences, University of Arizona
Tucson, AZ, USA
baltrus@email.arizona.edu

ABSTRACT

Horizontal gene transfer (HGT) is one of the most important evolutionary forces within microbial populations. While there has been extensive discussion over the beneficial effects of HGT, identifying and exploring fitness costs associated with DNA acquisition are essential for structuring general discussions about the interplay between selection and HGT. Research across a variety of systems has illuminated the evolutionary potential of these costs and these results can be used as a baseline to evaluate the dependence of costs on genomic and environmental contexts. Using such costs as the foundation for future studies will further evolutionary exploration the interface between acquired regions and recipient genomes and enable experimental evaluation of the role of HGT in structuring genetic diversity across populations.

Introduction

Horizontal gene transfer (HGT), the movement of genetic material across strain and species boundaries, is one of the most powerful yet least understood evolutionary forces. HGT drives evolutionary patterns within bacteria, but is not limited to microbes as gene transfer has been implicated in ecological and genomic transitions across eukaryotes [1, 2]. Despite an increasingly firmer grasp of the frequency, mechanisms, and phylogenetic implications of gene transfer, general evolutionary models have largely failed to incorporate this process [1] (although there are exceptions, i.e. [3]). It is possible that focusing on understanding only the beneficial effects of HGT limits generality of such results because of a lack of predictable outcomes, context specificity, and potential to overturn niche space and the adaptive landscape. While fitness costs associated with HGT may prove more applicable to general evolutionary models, to date there has been minimal exploration of these costs in and of themselves. Furthermore, amelioration of these costs through selection can leave lasting signatures of adaptation within these genomes and change future evolutionary dynamics and trajectories within populations. The goals for this manuscript are therefore to condense experimental results from a variety of research sources into a tangible framework, to highlight the potential for costs within and across species, and describe ways in which genomic and environmental context mediate these deleterious effects. The time is ripe to build a solid foundation for exploring the generality of fitness costs associated with HGT, which will

enable greater understanding of the evolutionary interplay between selection and gene transfer.

Evidence of a Fitness Cost for HGT

Early reports of direct costs for HGT center on the movement of engineered plasmids into *Escherichia coli* to carry out industrial production of recombinant proteins and plasmids [4, 5]. Transfer of plasmids into naive cells often altered growth properties across strains in such a general way that it warranted coining the phrase “metabolic burden” [4]. Reports detailed the lengthening of lag phase, lowering of maximum growth rates or cell densities, and changes in proxy measurements such as oxygen consumption across bacterial and eukaryotic systems [4, 6]. In parallel, direct measurements of the fitness consequences of HGT were undertaken as the research community sought ways to minimize the risks of antibiotic resistance, and specifically the hope that treatments could be modified to take advantage of deleterious fitness effects associated with the transfer of resistance factors [7]. To this end, multiple groups measured the cost of plasmid maintenance through comparisons of growth dynamics and head to head competition, with results largely mirroring those from industrial applications [8-10]. Additional reports have established a library of phenotypic effects that occur as byproducts of HGT including alterations in biofilm formation and thermal tolerance [11, 12]. On the other hand, and somewhat surprisingly, whole bacterial genomes have been incorporated into other organisms without large-scale fitness

consequences, although the search space for finding detrimental effects was limited [13, 14].

The rise of whole genome sequencing and comparative genomics has further enabled retrospective approaches to investigate the fitness outcomes of HGT. Although a majority of genes can be successfully transferred to *E. coli* under laboratory conditions, biases exist in the types of genes that undergo HGT and are maintained within genomes over extended time scales [15-17]. These biases have been interpreted as evidence of selection against the transfer of certain proteins, but one must be cautious because unrelated population level forces, like genetic drift, could also skew the potential for fixation. Even so, some genes remain recalcitrant to transfer even under optimal conditions [15], which implies that fitness barriers to HGT do exist for certain combinations of loci and genomic backgrounds. Taken together with the direct measurements from a wide variety of systems it is clear that, all else equal, any given HGT event carries with it significant potential for fitness costs. Despite these results, a clear consensus about the underlying genetic basis of these effects has failed to emerge and inherent correlations muddy definitive experimental tests.

Possible Causes of the Cost of HGT

Disruption of Genomic Features

Incorporation of transferred DNA into the chromosome inherently requires the disruption of chromosomal regions with randomness of integration sites depending on the mechanism of transfer [18]. Indeed, one of the early hallmarks associated with genomic islands was their close proximity to tRNA loci thought to act as target sites for integration [19]. Although the magnitude of the cost differs from gene to gene, disruption of an ecologically relevant gene as a byproduct of chromosomal integration will clearly alter phenotypes [20] and could lower fitness under specific conditions [21]. More general yet subtle costs could also arise if symmetry around the origin of replication of circular chromosomes is disrupted by integration of foreign DNA. Breaks in symmetry due to inversions or insertions have been shown to disrupt chromosomal replication and daughter cell formation, and are strongly selected against across bacterial species [22]. While many transferred regions may be small enough or far enough away form the origin to avoid disrupting symmetry, and genomes often display positional preferences for incorporating foreign DNA [23], the likelihood and overall importance of this cost remains unexplored in the context of HGT.

Sequestration of Limited Resources

The larger the genome the greater the baseline need for metabolic building blocks such as carbon, nitrogen, and phosphorous as well as consumption of molecular fuel (e.g. ATP) to carry out basic cellular processes. The precision and power of selection to act on metabolic efficiency in bacterial genomes is evidenced by metabolic

systems in *E. coli* that have evolved to minimize energetic cost under starvation [24-26]. Despite an often assumed cost for maintaining extra DNA [27], theoretical and experimental studies have demonstrated that this cost is predominantly due to transcription and translation [28, 29]. While extra DNA likely isn't costly to maintain due to replication limitations, the energetic bill quickly accrues with protein production because thousands of transcripts and proteins can be made from a single locus.

Shortages of molecular building blocks aren't the only physiological bottleneck during cell growth because the process of protein production is itself costly [29]. While ribosomes may be among the most abundant molecules within a cell, they also constitute a finite resource that directly impact almost every cellular process dependent on protein production. Ribosomal concentrations can limit protein production when in short supply, such as during the shift from nutritionally poor to rich environments, and under these conditions every additional translated gene could be selected against to avoid lowering the production capacity for essential proteins [29]. Similar arguments and costs for HGT can be extended to any pathways where extra DNA diverts limiting resources from more important uses, such as RNA polymerase occupancy during transcription.

Disruptive Interactions with Cellular Networks

The term "complexity hypothesis" was coined to describe the first hints of bias in HGT frequency across bacterial genomes because genes involved in complex cellular

processes, such as translation, were underrepresented among transferred regions [17]. Although subsequent studies have shown that multiple variables, such as gene expression, are also correlated with these biases, recent results point towards connectivity of proteins as the most direct determinant of these trends [16, 17]. Framed within the context of gene networks, the more hub-like a protein the less likely it is to undergo HGT.

Explanations for why protein connectivity influences HGT largely focus on disruption of cellular processes. Genes undergoing HGT have not had the chance to co-evolve with other networks and pathways within the recipient genome. It is therefore possible that divergent proteins function less efficiently within the context of recipient genetic networks than in the donor genome, with the magnitude of these deleterious effects growing with number of interactions [30]. This problem intensifies if the recipient genome lacks appropriate interaction partners for the transferred proteins, which may be related to the observation that HGT is more successful when proteins from the same pathways are already present within the cell [31]. Incorporation of extra homologues within a genome can also disrupt fine-tuning of cell physiology through protein dosage changes, which alter metabolic flux, especially if other members of the pathway were not transferred [16, 31, 32]. Alternatively, highly connected proteins or proteins with high levels of intrinsic disorder may be more likely to form spurious new connections within novel genomic contexts [16, 32, 33]. In this case, the cost could arise through collateral damage due to disruption of networks unrelated to the transferred proteins function. Despite extensive evidence from genome scale data, real time tests of the complexity

hypotheses have failed to clearly demonstrate the presence of detrimental fitness effects [30].

Cytotoxic Effects

Misfolded proteins lead to direct fitness costs within microbial cells through cytotoxic effects and the disruption of cellular processes [34]. Since HGT can place proteins within dramatically different cellular contexts than they have previously experienced or are optimized for, transferred proteins may be more prone to misfolding and inclusion body formation than those that have co-evolved with the recipient genome [35]. Misfolded proteins are normally dealt with through the action of proteases, export out of the cell, and chaperone binding, but in high enough levels can trigger lethal cellular stress due to membrane depolarization and OH- production, a response that resembles apoptosis [35, 36]. Recipient genomes may also lack correct suites of chaperone proteins to aid folding of foreign proteins thereby hastening cytotoxic effects [37].

Systems Level Effects

Regulatory networks within cells have evolved to respond to precise feedback loops that monitor transcriptional patterns and the concentrations of signaling substrates. General cell-wide costs may therefore arise as indirect effects of the horizontally transferred regions on metabolic flux and cellular processes or through

192 rewiring of cellular regulatory networks. For instance, one of the hallmarks of
193 recombinant protein production in *E. coli* is a shift in carbon flux [5]. Replication of
194 additional DNA through HGT can lead to a paucity of nucleotides within microbial cells,
195 switching metabolic networks towards energy production through the pentose
196 phosphate rather than the TCA cycle [5]. These shifts lead to the buildup of toxic
197 metabolic intermediates such as acetate, thereby lowering growth rates. Along these
198 same lines, acquisition of genes for which there is already a functioning copy within the
199 genome could disrupt the efficiency and throughput of metabolic flux within a cell
200 through changes in protein dosage [16, 38]. Indeed, the unifying theme that Sorek et al
201 found when comparing genes recalcitrant to transfer was that they were always present
202 as single copies within their original genomes [15].

203 Transferred regions may trigger expression of global regulatory circuits by
204 altering concentrations of key signaling molecules such as cAMP [5]. These effects
205 could occur through direct binding and sequestration of the signaling molecules or as
206 byproducts of changing the cellular environment. Likewise, misfolded proteins can
207 trigger the heat shock response by sequestering the chaperone DnaJ or the stringent
208 response through amino acid shortages [35, 39]. HGT dependent changes in global
209 regulation can trickle down through cellular networks to shift a wide variety of pathways
210 and causing phenotypic changes as collateral damage. Transferred regions may also
211 code for regulator proteins that interact with many chromosomal loci to alter transcript
212 levels throughout the genome [40]. Levels of cross-talk between transferred regions and
213 the background genome have not been systematically investigated across many

systems, but could directly affect fitness by altering ecologically relevant phenotypic responses such as iron acquisition [41] . All in all, large scale phenotypic changes and direct metabolic costs may occur at significant levels as transcriptional by-products of HGT.

Interactions Between DNA Sequence and Costs

HGTed regions have been identified by sequence specific signatures, like GC%, that differ from the rest of the recipient genome. These differences can profoundly alter selection pressures against foreign DNA, magnifying any costs already present. For instance, amino acids made from AT rich codons are less energy efficient to synthesize and their transfer could be more strongly selected against within nutritionally poor environments [42]. Alternatively, regions may be differentially expressed upon transfer to recipient genomes due to the action of global regulators such as H-NS [43, 44], thereby exacerbating other underlying costs (Box 1). The fitness effects of GC content were recently tested in *E. coli* using GFP and ϕ29 DNA polymerase constructs engineered to differ in GC% but not in protein sequence or codon bias [45]. Although the mechanistic basis of the result remains unclear, growth rates were inversely correlated with GC content of the foreign loci.

Natural selection molds codon usage so that each genome maintains a specific codon bias [46]. Inefficient codon patterns within the acquired region can also

exacerbate other general fitness effects of HGT [37]. Evidence for a direct link between metabolic burden and coding inefficiency includes the finding that engineered proteins utilizing rare codons lead to higher fitness costs than those using abundant codons [39]. This cost was attributed to triggering of the stringent response due to lack of charged amino acids. Likewise, mismatches between amino acid usage and codon bias lower translational accuracy and efficiency [37]. At a baseline, the translational mutation rate is substantially higher than genomic or transcriptional mutation rates and improper codon usage only increases errors. Higher mutation rates during translation lead to an increase in misfolded proteins and could thereby skew the chances of cytotoxic effects for recently acquired genes. Improper codon usage will also cause ribosomes to stall as they translate regions of HGT [46], potentially adding to the detrimental effects of ribosome sequestration. Direct experimental tests of fitness and codon bias are rare, but Kudla et al. demonstrated that overexpression of a gene with incorrect codon bias is deleterious while Tuller et al. showed that HGT was more likely to occur between genomes that have similar codon biases [47, 48]. All of the evidence points to sequence specific effects influencing the magnitude of HGT costs in a generalized fashion no matter what the underlying basis of costs.

Strain and Environment Effects on the Cost of HGT

The cost of HGT depends on the strain, the environment encountered, and the mechanism of transfer (Box 2). At a broad level, some lineages and species will be

more sensitive to costs of HGT due to skewed codon usage or ribosome levels. Larger genomes tend to be more receptive to gene transfer, which may indicate that tolerance to the costs of HGT correlates with genome size but could also be explained through ecological similarity [49]. It is certainly possible that bacteria with larger genomes have optimized replication and repair systems to deal with extra DNA, and that the cost of extra genes therefore decreases as a percentage of total genomic content. Even within a species there have been a variety of reports showing strain specificity for phenotypes related to HGT and fitness effects, such as plasmid copy number or stability [50, 51]. Further evidence for strain specific effects comes from studies focused on engineering bacterial strains with lower metabolic burdens for manufacturing purposes [5]. One of the ways to counter metabolic burden is through the addition of extra genes or pathway connections that rewire metabolic flux to prevent the build up of toxic intermediates. If such changes are possible in the lab, it is straightforward to imagine that natural polymorphisms exist within these same genes and pathways and that this variation could make some strains more tolerant of costs and receptive to HGT. Extending these ideas to the transcriptional level, polymorphisms across strains in global regulators like HN-S likely affect overall levels of gene expression for transferred regions across strain backgrounds [43, 52].

The magnitude of fitness effects for any evolutionary event, including HGT, depends strongly on how selection pressures are structured by the environment [53]. A clear demonstration of environmentally dependent costs involves chromosomal and gene disruptions at loci which are necessary only under certain ecological conditions

[54]. As examples, the fitness effects of a variety of horizontally acquired antibiotic resistance genes has been shown to differ *in vitro* and *in vivo [55]* and HGT could be more costly during *E. coli* growth in glucose culture rather than glycerol because of the buildup of acetate [5]. Environments also vary greatly in their nutritional quality. One of the most striking trends in bacterial genomics is that marine bacteria possess streamlined genomes with very little extraneous and non-coding DNA [56]. While there is debate about the root cause of genome streamlining, some have argued that low levels of nutrients in the ocean, and specifically nitrogen, directly selects for smaller genomes [57]. Although there have been no general studies on the overall levels of HGT within these lineages compared to terrestrial systems, one can imagine that energetic costs of acquiring extra DNA are magnified in marine environments and other nutrient poor systems.

Conclusions

Although there have been comparatively few direct tests of the costs of HGT, studies from a variety of research fields suggest that every transfer event carries potential fitness costs. While these costs are offset in cases where HGT events provide benefits, their existence leaves lasting genomic signatures and can alter future evolutionary trajectories. These detrimental effects can be general or nuanced, but in all cases depend greatly on interactions between genomic and environmental contexts. These costs are not static, and can be minimized over time by selection as the

302 transferred regions are molded to fit within each genome and each genome is in turn
303 altered to accommodate foreign DNA. As the intricacies of deleterious effects across
304 HGT events are further flushed out, they can be incorporated into evolutionary models
305 of microbial populations, used to establish further laboratory experiments to validate and
306 quantify hypotheses underlying the costs, and provide a general framework to illuminate
307 the role of HGT across populations.
308
309

BOX 1

There is a tendency for newly acquired operons to be expressed at suboptimal levels [58]. Nearly all of the costs of HGT will be exacerbated by increases in gene expression of the acquired region [16]. For instance, higher levels of gene expression require additional resources both in energy and in ribonucleotide or amino acid pools [28]. The greater the number of transcripts, the more likely are the deleterious effects of ribosome occupancy and sequestration. Fitness costs associated with protein interactions and cytotoxicity increase proportionally with higher protein concentrations [33, 59]. In sum, one of the few generalities evident thus far across all HGT events is that the potential and magnitude for fitness costs increases with gene expression levels.

Given the prevalence of HGT within microbial communities, it should be no surprise that cells have evolved ways to minimize costs associated with over or improper expression. One method to lower the cost of rampant over-expression of transferred regions involves HN-S proteins, orthologs, and related systems [43]. HN-S proteins are found throughout bacterial species and generally act by binding regions of high AT content, which can either silence or promote expression based on GC% and genomic context. Interestingly, the co-evolutionary importance of HN-S like proteins in defending against HGT is reinforced by the demonstration that genomic parasites such as phage and plasmids harbor anti-HN-S proteins in order to maintain transcription in the face of genomic repression [43, 60]. Similar scenarios play out with regard to other regulatory systems, as phage proteins have been shown to manipulate the heat shock

response and some plasmids have evolved to dampen the bacterial SOS response after movement into a naive cell [61, 62].

One of the most interesting forms of expression modulation involves an episome containing virulence genes within the bean pathogen *Pseudomonas syringae* pv. *phaseolicola,* which triggers immune recognition in some plant hosts when expressed [54]. This episome can freely replicate outside of the genome but can also integrate into specific genomic locations and is passed across strains through natural transformation. However, *P. syringae* can avoid recognition through extra-genomic silencing based on gene silencing of this transferred region when the episome excises itself from the chromosome.

Box 2

The potential for costs associated with HGT depends on the mechanisms of transfer. If chromosomal regions are simply replaced through recombination, either through natural transformation or other processes, fitness costs are likely to involve inefficient protein functions and resemble the deleterious effects sex due to the breakup of co-evolved alleles [63]. In situations where DNA is added to the genome, multiple classes of costs can arise based on the size of the vectors involved. Small phage and plasmids don't harbor many loci other than those necessary for their own replication and the costs are likely to be enriched for gene disruptions, energetic requirements, and specific protein level interactions as often seen with antibiotic resistance loci [10]. Smaller plasmids in particular may be more prone to energetic costs because of an inverse relationship between copy number and plasmid size [64]. In conjugative plasmids, significant amounts of resources may be shunted towards the production of the transfer apparatus and metabolic costs are potentially skewed so that they are correlated with transfer rates to naive cells [65].

Nutritional requirements obviously increases with the size of the vector involved, but larger phage and plasmids add layers of complexity due to secondary genes carried as passengers [66]. Phage can carry divergent copies of housekeeping loci, such as DNA polymerases, that can interfere with or modify basic cellular processes and evolutionary relevant parameters such as mutation rates [67]. Perhaps the most interesting case involves megaplasmids (or chromids), which are large secondary

366 genomic elements found in 1 out of 10 bacterial species [68]. Some of these large
367 chromids are conjugative and can transfer readily from cell to cell [69]. Megaplasmids
368 often contain a wide variety of housekeeping genes that may only be slightly divergent
369 from the chromosomal copy. For instance, pMPPla107, a megaplasmid within *P.*
370 *syringae*, potentially contains a ribosomal protein, 30 tRNA loci in almost the same
371 proportions as is present on the main chromosome, as well as roughly 50 genes
372 involved in basic pathways such as nucleotide metabolism and DNA replication [70].
373 However, none of these pathways is complete so it is possible that complete function
374 requires interactions between chromosomal loci and proteins from the megaplasmid.
375 Although just an observation at this point, strains that naturally harbor pMPPla107 grow
376 more slowly both *in vitro* and *in planta.* The larger the vector for HGT, the greater the
377 opportunities to disrupt basic cellular processes through detrimental protein interactions.
378

1 Syvanen, M. (2012) Evolutionary implications of horizontal gene transfer. *Annu Rev Genet* 46, 341-358
2 Keeling, P.J. and Palmer, J.D. (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 9, 605-618
3 Mozhayskiy, V. and Tagkopoulos, I. (2012) Horizontal gene transfer dynamics and distribution of fitness effects during microbial *in silico* evolution. *BMC Bioinformatics* 13, S13
4 Diaz-Ricci, J.C. and Hernández, M.E. (2000) Plasmid effects on *Escherichia coli* metabolism. *Crit Rev Biotechnol* 20, 79-108
5 Gonçalves, G.A.*, et al.* (2011) Rational engineering of *Escherichia coli* strains for plasmid biopharmaceutical manufacturing. *Biotechnol J* 7, 251-261
6 Rensburg, E.*, et al.* (2012) The metabolic burden of cellulase expression by recombinant *Saccharomyces cerevisiae* Y294 in aerobic batch culture. *Appl Microbiol Biotechnol* 96, 197-209
7 Svara, F. and Rankin, D.J. (2011) The evolution of plasmid-carried antibiotic resistance. *BMC Evol Biol* 11, 130
8 Heuer, H.*, et al.* (2007) Frequent conjugative transfer accelerates adaptation of a broad-host-range plasmid to an unfavorable *Pseudomonas putida* host. *FEMS Microbiol Ecol* 59, 738-748
9 Subbiah, M.*, et al.* (2011) Selection Pressure Required for Long-Term Persistence of blaCMY-2-Positive IncA/C Plasmids. 77, 4486-4493
10 Harrison, E. and Brockhurst, M.A. (2012) Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends Microbiol* 20, 262-267
11 Gaillard, M.*, et al.* (2008) Host and invader impact of transfer of the *clc* genomic island into *Pseudomonas aeruginosa* PAO1. *Proc Natl Acad Sci USA* 105, 7058-7063
12 Sato, T. and Kuramitsu, H. (1998) Plasmid maintenance renders bacteria more susceptible to heat stress. *Microbiol Immunol* 42, 467-469
13 Tagwerker, C.*, et al.* (2012) Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic Acids Res* 40, 10375-10383
14 Itaya, M.*, et al.* (2005) Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc Natl Acad Sci USA* 102, 15971-15976
15 Sorek, R.*, et al.* (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318, 1449-1452
16 Park, C. and Zhang, J. (2012) High expression hampers horizontal gene transfer. *Genome Biol Evol* 4, 523-532
17 Cohen, O.*, et al.* (2011) The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol* 28, 1481-1489
18 Green, B.*, et al.* (2012) Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mob DNA* 3, 3

19 Fernandez-Gomez, B.*, et al.* (2012) Patterns and architecture of genomic islands in marine bacteria. *BMC Genomics* 13, 1-1
20 Rabinovich, L.*, et al.* (2012) Prophage excision activates *Listeria* competence genes that promote phagosomal escape and virulence. *Cell* 150, 792-802
21 Hayes, F. (2003) Transposon based strategies for microbial functional genomics and proteomics *Annu Rev Genet* 37, 3-29
22 Darling, A.E.*, et al.* (2008) Dynamics of genome rearrangement in bacterial populations. *PLoS Genet* 4, e1000128
23 Touchon, M.*, et al.* (2009) Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genet* 5, e1000344
24 Bragg, J.G.*, et al.* (2012) Protein elemental sparing and codon usage bias are correlated among bacteria. *Mol Ecol* 21, 2480-2487
25 Vilaprinyo, E.*, et al.* (2010) Minimization of biosynthetic costs in adaptive gene expression responses of yeast to environmental changes. *PLoS Comput Biol* 6, e1000674
26 Smith, D. and Chapman, M. (2010) Economical evolution: microbes reduce the synthetic cost of extracellular proteins. *MBio* 1, e00131-00110-e00131-00118
27 Raffaele, S. and Kamoun, S. (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol* 10, 417-430
28 Bragg, J.G. and Wagner, A. (2009) Protein material costs: single atoms can make an evolutionary difference. *Trends Genet* 25, 5-8
29 Shachrai, I.*, et al.* (2010) Cost of unneeded proteins in E. coli is reduced after several generations in exponential growth. *Mol Cell* 38, 758-767
30 Wellner, A. and Gophna, U. (2008) Neutrality of foreign complex subunits in an experimental model of lateral gene transfer. *Mol Biol Evol* 25, 1835-1840
31 Pal, C.*, et al.* (2005) Horizontal gene transfer depends on gene content of the host. *Bioinformatics* 21, ii222-ii223
32 Vavouri, T.*, et al.* (2009) Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* 138, 198-208
33 Yang, J.R.*, et al.* (2012) Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc Natl Acad Sci USA* 109, E831-840
34 Geiler-Samerotte, K.A.*, et al.* (2011) Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proc Natl Acad Sci USA* 108, 680-685
35 Jürgen, B.*, et al.* (2010) Quality control of inclusion bodies in *Escherichia coli*. *MIcrob Cell Fact* 9, 41
36 Kohanski, M.A.*, et al.* (2008) Mistranslation of membrane proteins and two-component system activation trigger antibiotic-mediated cell death. *Cell* 135, 679-690
37 Drummond, D.A. and Wilke, C.O. (2009) The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet* 10, 715-724
38 Gout, J.F.*, et al.* (2010) The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet* 6, e1000944
39 Bonomo, J. and Gill, R.T. (2005) Amino acid content of recombinant proteins influences the metabolic burden response. *Biotechnol Bioeng* 90, 116-126

40 Chen, Y.*, et al.* (2005) Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. *PLoS Biol* 3, e229
41 Shintani, M.*, et al.* (2009) Response of the *Pseudomonas* host chromosomal transcriptome to carriage of the IncP-7 plasmid pCAR1. *Environ Microbiol*
42 Akashi, H. and Gojobori, T. (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci USA* 99, 3695-3700
43 Doyle, M.*, et al.* (2007) An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science* 315, 251-252
44 Baños, R.C.*, et al.* (2009) Differential regulation of horizontally acquired and core genome genes by the bacterial modulator H-NS. *Proc Natl Acad Sci USA* 5, e1000513
45 Raghavan, R.*, et al.* (2012) A selective force favoring increased G+C content in bacterial genes. *Proc Natl Acad Sci USA* 109, 14504-14507
46 Gingold, H. and Pilpel, Y. (2011) Determinants of translation efficiency and accuracy. *Mol Syst Biol* 7, 1-13
47 Tuller, T.*, et al.* (2011) Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Res* 39, 4743-4755
48 Kudla, G.*, et al.* (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324, 255-258
49 Cordero, O.X. and Hogeweg, P. (2009) The impact of long-distance horizontal gene transfer on prokaryotic genome size. *Proc Natl Acad Sci USA* 106, 21748-21753
50 Humphrey, B.*, et al.* (2012) Fitness of *Escherichia coli* strains carrying expressed and partially silent IncN and IncP1 plasmids. *BMC Microbiol* 12, 53
51 Singer, A.*, et al.* (2009) DNA plasmid production in different host strains of *Escherichia coli*. *J Ind Microbiol* 36, 521-530
52 Miyakoshi, M.*, et al.* (2009) High-resolution mapping of plasmid transcriptomes in different host bacteria. *BMC Genomics* 10, 12
53 Slater, F.R.*, et al.* Progress towards understanding the fate of plasmids in bacterial communities. *FEMS Microbiol Ecol* 66, 3-13
54 Godfrey, S.A.*, et al.* (2011) The stealth episome: suppression of gene expression on the excised genomic island PPHGI-1 from *Pseudomonas syringae* pv*. phaseolicola*. *PLoS Pathog* 7, e1002010
55 Enne, V.*, et al.* (2005) Assessment of the fitness impacts on *Escherichia coli* of acquisition of antibiotic resistance genes encoded by different types of genetic element. *J Antimicrob Chemother* 56, 544-551
56 Grote, J.*, et al.* (2012) Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *MBio* 3, e00252-00212-e00252-00212
57 Grzymski, J.J. and Dussaq, A.M. (2011) The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J* 6, 71-80
58 Chou, H.H.*, et al.* (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332, 1190-1192
59 Drummond, D.A. and Wilke, C.O. (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution *Cell* 134, 341-352

60 Skennerton, C.T.*, et al.* (2011) Phage encoded H-NS: a potential achilles heel in the bacterial defence system. *PLoS One* 6, e20095
61 Petrova, V.*, et al.* (2009) An SOS inhibitor that binds to free RecA protein: the PsiB protein. *Mol Cell* 36, 121-130
62 Perrody, E.*, et al.* (2012) A bacteriophage-encoded J-domain protein interacts with the DnaK/Hsp70 chaperone and stabilizes the heat-shock factor σ32 of *Escherichia coli*. *PLoS Genet* 8, e1003037
63 Otto, S.P. (2009) The evolutionary enigma of sex. *Am Nat* 174 Suppl 1, S1-S14
64 Zünd, P. and Lebek, G. (1980) Generation time-prolonging R plasmids: correlation between increases in the generation time of *Escherichia coli* caused by R plasmids and their molecular size. *Plasmid* 3, 65-69
65 Turner, P.E.*, et al.* (1998) Tradeoff between horizontal and vertical modes of transmission in bacterial plasmids. *Evolution*, 315-329
66 Hatfull, G.F. (2008) Bacteriophage genomics. *Curr Opin Micribol* 11, 447-453
67 Liu, J.*, et al.* (2006) Protein repertoire of double-stranded DNA bacteriophages. *Virus Res* 117, 68-80
68 Harrison, P.W.*, et al.* (2010) Introducing the bacterial "chromid": not a chromosome, not a plasmid. *Trends Microbiol*
69 Smillie, C.*, et al.* (2010) Mobility of plasmids. *Microbiol Mol Biol Rev* 74, 434-452
70 Baltrus, D.A.*, et al.* (2011) Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae i*solates. *PLoS Pathog* 7, e1002132