Review

Cell PRESS

Exploring the costs of horizontal gene transfer

David A. Baltrus

School of Plant Sciences, University of Arizona, Tucson, AZ 85721-0036, USA

Horizontal gene transfer (HGT) is one of the most important evolutionary forces within microbial populations. Although evidence for beneficial fitness effects of HGT is overwhelming, recently acquired regions often function inefficiently within new genomic backgrounds so that each transfer event has the potential to disrupt existing regulatory and physiological networks. Identifying and exploring costs is essential for guiding general discussions about the interplay between selection and HGT, as well as generating hypotheses to explain how HGT affects evolutionary potential through, for example, changing adaptive trajectories. Focusing on costs of HGT as foundations for future studies will enhance exploration at the interface between acquired regions and recipient genomes, including the process of amelioration, and enable experimental evaluation of the role of HGT in structuring genetic diversity across populations.

HGT: *caveat emptor*

HGT, the movement of genetic material across strain and species boundaries, is one of the most powerful yet least understood evolutionary forces. The power of HGT to transform evolutionary landscapes has become increasingly evident with the number of sequenced genomes, even within eukaryotes, with fitness effects that range from fine-tuning of enzymatic function to enabling major ecological transitions [1–7]. To be maintained within a population, acquired regions must provide a great enough benefit or low enough fitness cost to avoid loss due to genetic drift or selection. However, recently acquired regions often function inefficiently within their new genomic backgrounds so that, despite great evolutionary benefits, they can be energetically or physiologically costly [8,9]. Furthermore, every transfer event carries with it the potential to generate detrimental side effects, which become especially relevant in cases where the regions of HGT create the sole selection pressures for maintenance (i.e., addiction modules [10]). Although energetic waste and side effects can influence future evolutionary dynamics and adaptive trajectories, to date there has been minimal exploration of the costs associated with HGT in and of themselves.

I define costs of HGT as inefficiencies that move genotypes down an adaptive slope, and also include phenotypic changes that are not directly selected in the focal environment but which can alter evolutionary potential across environments. Below, I condense experimental results from a variety of research sources into a tangible framework that enables exploration of these costs, highlighting the potential for costs within and across species, and pointing towards ways in which genomic and environmental context influence the strength of selection. Although HGT can also lead to gene conversion or replacement of whole pathways, for simplification I focus only on transfer events where 'extra' DNA is incorporated into genomes. The time is ripe to build a solid foundation for exploring the generality and plasticity of fitness costs associated with HGT to facilitate studies focused on evolutionary interplay as transferred regions integrate into recipient genomes.

Evidence of phenotypic costs

Early reports of direct costs for HGT center on the movement of engineered plasmids into *Escherichia coli* for the purpose of recombinant protein and plasmid production [11,12]. Transfer of plasmids into naive cells often altered growth properties across strains in such general ways that it warranted coining of the phrase 'metabolic burden' [11]. 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 [11,13]. Additionally, multiple groups have measured the cost of plasmid maintenance through comparisons of growth dynamics and head-to-head competition, with results largely mirroring those from industrial applications [14–16]. Other studies have catalogued a library of phenotypic effects that, potentially, are byproducts of HGT, including alterations in biofilm formation and thermal tolerance [17,18]. By contrast, 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 [19,20].

The rise of whole-genome sequencing and comparative genomics has further enabled retrospective approaches to investigate the fitness outcomes of HGT. Although most genes appear to be successfully transferred in nature and under laboratory conditions [21,22], biases exist in the types of genes and pathways that undergo HGT and are maintained within genomes over extended timescales [8,23,24]. Moreover, some genes remain recalcitrant to transfer even under optimal laboratory conditions [21], which implies that fitness barriers to HGT do exist for certain combinations of loci and genomic backgrounds. Taken together, evidence from a wide variety of systems demonstrates that, all else being equal, any given HGT

Corresponding author: Baltrus, D.A. (baltrus@email.arizona.edu).

0169-5347/$ – see front matter
 http://dx.doi.org/10.1016/j.tree.2013.04.002

event carries with it significant potential for costs. Despite these results, a clear consensus about the underlying genetic basis of these effects has failed to emerge.

Mechanisms behind the cost of HGT

To facilitate discussion, I have split the mechanisms underlying costs into separate categories described below and illustrated in Figure 1. However, it is important to recognize that these underlying mechanisms are not mutually exclusive, especially those involving transcriptional and translational responses.

Disruption of genomic features

Incorporation of transferred DNA into the chromosome inherently requires the disruption of chromosomal regions, with site specificity for these recombination events determined by the mechanism of transfer [25]. 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 [26] and could lower fitness under specific conditions [27]. 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. Although evidence hints that changes leading to asymmetry are deleterious, with explanations centered on disruption of transcriptional complexes or timing of replication, the precise selection pressures behind these patterns remain unclear [28,29].

Cytotoxic effects

Misfolded proteins lead to direct fitness costs within microbial cells through cytotoxic effects and the disruption of cellular processes [30]. Given that HGT can place proteins within dramatically different cellular contexts than they have previously experienced or are optimized for, transferred proteins might be more prone to misfolding than those that have coevolved with the recipient genome [31,32]. Recipient genomes might also lack correct suites of chaperone proteins to aid folding of foreign proteins, thereby hastening cytotoxic effects [33]. In high enough concentrations, misfolded proteins can trigger lethal cellular stress due to membrane depolarization and OH^- production [31,34].

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. Despite an often assumed cost for maintaining extra DNA, theoretical and experimental studies have demonstrated that this cost is predominantly due to transcription and translation [35,36]. Although extra DNA is unlikely to be 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 [35]. An additional, but related, cost also exists in that extra genomic regions sequester cellular machinery from performing critical housekeeping processes [36]. Given that ribosomal concentrations limit protein production when in short supply, every additional translated gene can be selected against by lowering the production capacity for essential proteins. Similar sequestration arguments can be extended to include 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 [23]. Although subsequent studies have shown that multiple variables, such as gene expression, are also correlated with these biases, recent results point towards the connectivity of proteins as the most direct determinant of these trends [8,23]. Explanations for why protein connectivity influences HGT largely focus on disruption of cellular processes. Genes undergoing HGT have not had the chance to coevolve with other networks and pathways within the recipient genome, and it is possible that divergent proteins function less efficiently within the context of recipient genetic networks than they do in the donor genome [37]. This problem intensifies with the number of interactions as well as if the recipient genome lacks appropriate interaction partners for the transferred proteins [38]. Incorporation of extra homologs within a genome can disrupt the fine-tuning of cell physiology through protein dosage changes, which alter metabolic flux, especially if other members of the pathway were not transferred [8,38,39]. Alternatively, highly connected proteins or those with high levels of intrinsic disorder might be more likely to form spurious new connections within novel genomic contexts [8,39,40].

System-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 can therefore arise as indirect effects of the horizontally transferred regions on metabolic flux and cellular processes or through rewiring of cellular regulatory networks. For instance, replication of additional DNA can lead to a paucity of nucleotides within microbial cells, which can cause a buildup of toxic metabolic intermediates, such as acetate, due to constraints on molecular pathway architecture [12]. Along these same lines, acquisition of genes for which there is already a functioning copy within the genome could disrupt the efficiency and throughput of metabolic flux through changes in protein dosage [8,41].

Transferred regions can trigger expression changes within global regulatory circuits by altering concentrations of key signaling molecules, such as cAMP [12]. These effects could occur through direct binding and sequestration of the signaling molecules or as byproducts of changing the cellular environment [31,42]. HGT-dependent changes in global regulation can trickle down through cellular networks to shift a wide variety of pathways and cause phenotypic changes as collateral damage. Transferred regions can also encode regulator proteins that directly

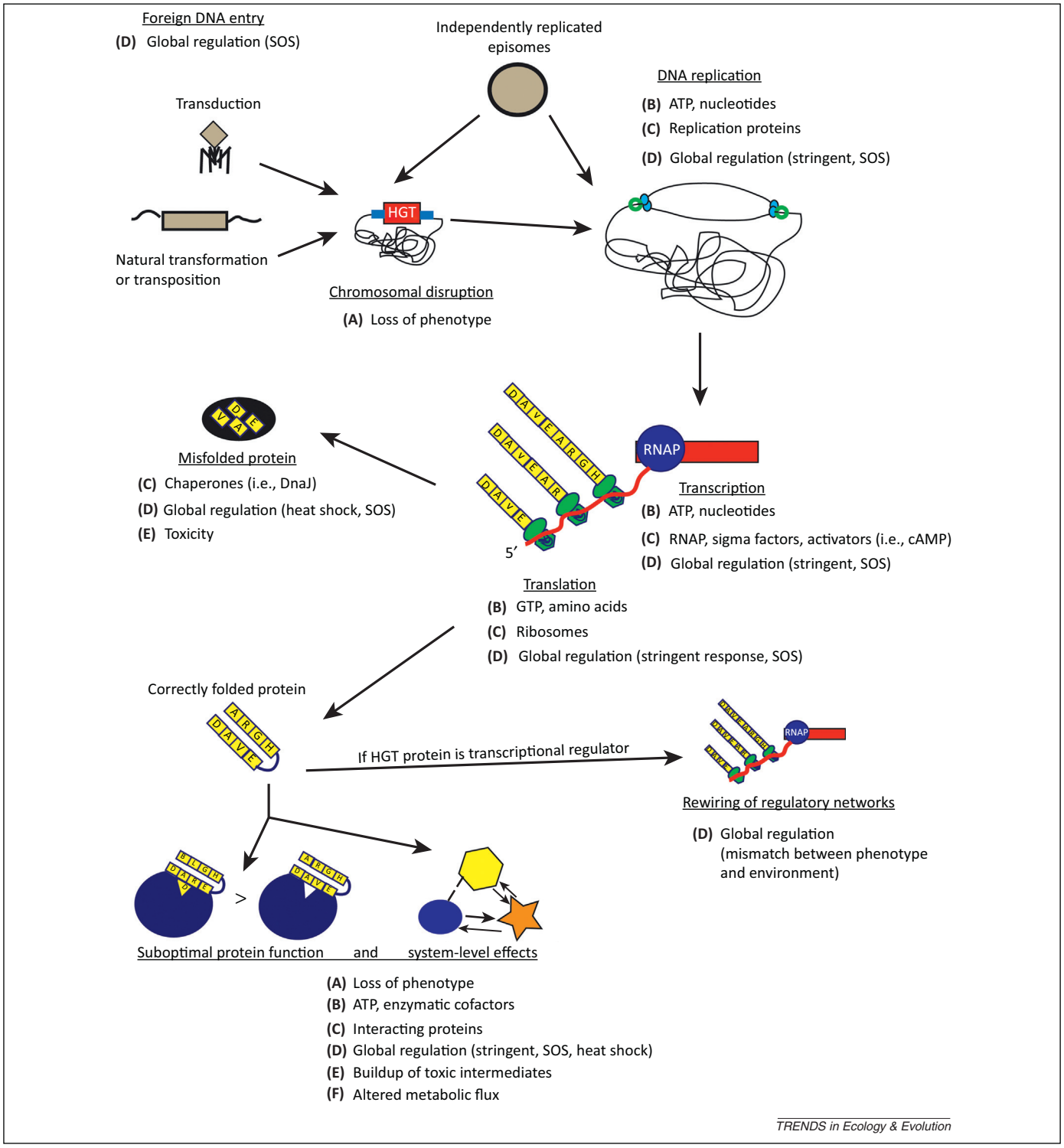

Figure 1. Mechanisms that create costs of horizontal gene transfer (HGT). Costs associated with recently acquired DNA regions are broken down into six classes, and the contributions of each class to various physiological and genetic processes are shown. **(A)** Phenotypic changes due to chromosomal disruption by horizontally acquired regions. If horizontally transferred genes are incorporated into the chromosome, they must disrupt existing genomic sequences, and such changes could be deleterious. **(B)** Energetic costs due to consumption of molecular building blocks or energy sources. Physiological processes require input of molecules such as nucleotides or ATP, and acquired regions can lower available amounts within the cell. **(C)** Sequestration of critical cellular processes. Fitness can be lowered if critical physiological processes require access to molecular machinery (i.e., ribosomes), but transferred regions actively occupy this machinery. **(D)** Changes in global regulatory circuits. Transcriptional responses can be sensitized by gene transfer so that they are triggered at lower amounts of stress than usual. Fitness costs can also occur when ecologically important regulatory circuits are rewired by crosstalk from horizontally acquired regulators. **(E)** Cytotoxicity. In high enough concentrations, misfolded proteins can lead directly to cell death. **(F)** System-level effects. Fitness can be lowered if horizontally acquired regions cause pathways to function suboptimally, either through inefficiency or disruption of molecular interactions.

affect transcript levels throughout the genome [43]. Levels of crosstalk between transferred regions and the background genome have not been investigated across many systems, but could directly affect fitness by altering ecologically relevant phenotypic responses. For instance, the presence of the plasmid pCARI significantly increased gene expression of a major siderophore in *Pseudomonas* strains [44]. Such a response could substantially alter

ecological relations of these strains because siderophores are the main pathway for iron acquisition in bacteria [45]. All in all, large-scale phenotypic changes and direct metabolic costs can occur at significant levels as transcriptional byproducts of HGT.

Interactions between DNA sequence and costs

One of the methods to identify transferred regions is through sequence-specific signatures, such as GC% and tetranucleotide bias, that differ from the rest of the recipient genome [46]. 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 [47]. Alternatively, regions can be differentially expressed upon transfer to recipient genomes due to the action of global regulators, such as histone-like nucleoid structuring protein (H-NS) [48], 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 [49]. Although the mechanistic basis of the result remains unclear, growth rates were inversely correlated with GC content of the foreign loci.

Box 1. Gene expression and costs of HGT

There is a tendency for newly acquired operons to be expressed at suboptimal levels [9]. Nearly all of the costs of HGT will be exacerbated by increases in gene expression of the acquired region [8]. For instance, higher levels of gene expression require additional resources both in energy and in ribonucleotide or amino acid pools [35]. 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 [40,75]. In sum, one of the few generalities evident thus far across all HGT events is that the potential and magnitude for fitness cost 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 overexpression of transferred regions involves HN-S proteins, orthologs, and related systems [48]. 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 coevolutionary importance of HN-S-like proteins in defending against HGT is reinforced by the demonstration that phage and plasmids harbor anti-HN-S proteins to maintain transcription in the face of genomic repression [76,77]. Similar scenarios play out with regard to other regulatory systems, because 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 [78,79].

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 [80]. This episome can freely replicate outside of the genome, integrate into specific genomic locations, and is passed across strains through natural transformation. However, *P. syringae* can avoid recognition through gene silencing of this transferred region when the episome excises itself from the chromosome.

Natural selection molds codon usage so that each genome maintains a specific codon bias [50]. As with changes in gene expression, inefficient codon patterns within the acquired region can also exacerbate other general fitness effects of HGT [33]. Evidence for a direct link between metabolic burden and coding inefficiency includes the finding that engineered proteins utilizing rare codons led to higher fitness costs than those using abundant codons [42], which 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 [33]. 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 [50], potentially adding to the detrimental effects of ribosome sequestration. Direct experimental tests of fitness and codon bias are rare, but Agashe *et al.* recently demonstrated that deleterious fitness effects of incorrect codon bias (surprisingly, both for loci engineered to contain either an over- or underabundance of common codons) arose from inefficient protein synthesis [51]. More relevant to studies of horizontal transfer, Tuller *et al.* computationally demonstrated that HGT was more likely to occur between genomes with similar codon biases [52]. 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. Although a correlation, larger genomes tend to be more receptive to gene transfers from divergent organisms [53], which might indicate that the effects of factors such as codon bias decrease with total genomic content. Even within a species, there has been a variety of reports showing strain specificity for phenotypes related to HGT and fitness effects, such as plasmid copy number or stability [54–56]. Moreover, 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 buildup of toxic intermediates [12]. If such changes are possible in the laboratory, 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 such as HN-S likely affect qualitative and quantitative levels of gene expression for transferred regions across strain backgrounds [48,57]. It is unclear how such strain-specific variation affects fitness after transfer, but it is certainly possible that regulatory or physiological interference lowers the overall success rate of HGT.

Box 2. Routes of HGT can influence costs

The potential for costs associated with HGT depends on the mechanisms of transfer. If chromosomal regions are simply replaced, either through natural transformation or other processes, fitness costs likely involve inefficient protein functions and resemble the separation of coevolved alleles through recombination [81]. In situations where DNA is added to the genome, multiple classes of cost can arise based on the size of the vectors involved. Small phage and plasmids do not harbor many loci other than those necessary for their own replication and the costs are potentially enriched for gene disruptions, energetic requirements, and specific protein level interactions, as often seen with antibiotic resistance loci [16]. Smaller plasmids in particular might be more prone to energetic costs because of an inverse relation between copy number and plasmid size [82]. In conjugative plasmids, significant amounts of resources might be shunted towards the production of the transfer apparatus and metabolic costs can be skewed so that they are correlated with transfer rates to naive cells [72].

Nutritional requirements increase with the size of the vector involved, but larger phage and plasmids add layers of complexity due to secondary genes carried as passengers [83]. 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 [84]. Perhaps the most interesting case involves megaplasmids (or chromids), which are large secondary genomic elements found in one out of ten bacterial species [85]. Some of these large chromids are conjugative and can transfer readily from cell to cell [86]. Megaplasmids often contain a wide variety of housekeeping genes that might only be slightly divergent from the chromosomal copy. For instance, pMPPla107, a megaplasmid within *Pseudomonas syringae*, potentially contains ribosomal proteins, tRNA loci in almost the same proportions as is present on the main chromosome, as well as approximately 50 genes involved in basic pathways, such as nucleotide metabolism and DNA replication [87]. However, none of these pathways appears complete, so it is possible that their function requires interactions between chromosomal loci and proteins from the megaplasmid. Although just an observation at this point, strains that naturally harbor pMPPla107 grow more slowly both *in vitro* and *in planta*. The larger the vector for HGT, the greater the opportunities to disrupt basic cellular processes through detrimental protein interactions.

Box 3. Outstanding questions and future directions

- How prevalent are the costs or side effects of HGT events?
- How many and what type of mutational steps allow full integration of acquired regions?
- How often do neutral or detrimental HGT events change future adaptive dynamics?
- Do environment and genomic background influence selection pressures on acquired regions in quantitative or qualitative ways?
- If multiple phenotypic changes occur as a side effect of HGT, are the causes of these changes independent or correlated?
- Can costs of HGT be exploited for therapeutic treatments?

The magnitude of fitness effects for any evolutionary event, including HGT, depends strongly on how selection pressures are structured by the environment [58]. For instance, the fitness effects of a variety of horizontally acquired antibiotic resistance genes have been shown to differ *in vitro* and *in vivo* [59]. Nutritional quality of the environment can also modify the magnitude of fitness costs, as has been shown for the *Agrobacterium* Ti plasmid [60]. One of the most striking trends in bacterial genomics is that marine bacteria have streamlined genomes with very little extraneous and noncoding DNA [61]. Although there is debate about the root cause of genome streamlining, some have argued that low levels of nutrients in the ocean, specifically nitrogen, directly select for smaller genomes [62]. Although there have been no general studies on the overall levels of HGT within these lineages compared with terrestrial systems, one can imagine that energetic costs of acquiring extra DNA are magnified in marine environments and other nutrient-poor systems (Box 3).

How side effects of HGT can shift the adaptive landscape

Every genotypic change can leave lasting signatures on evolutionary dynamics across populations [9,63,64]. If newly acquired regions are not immediately and efficiently integrated into new genomic contexts, the presence of HGT-related costs inherently increases the pool of beneficial mutations available within that population, altering both evolutionary rates and adaptive trajectories. Furthermore, because HGT facilitates ecological shifts, a general result might be that the first adaptive steps after microbes fill a new niche are those that optimize recently acquired regions or compensate for phenotypic side effects. Given enough time, it is clear that selection efficiently acts to ameliorate even the most subtle of costs [65–67], but what is the nature of these compensatory mutations and how might transfer events alter future evolutionary paths?

Prospective evolution experiments give some indication because they have shown that compensatory mutations to the costs of HGT can occur both inside and outside of the transferred region, and can involve interactions between foreign genes and the recipient genome [16]. Perhaps the most prevalent mutational change within these experiments involves deletion of specific genes and operons within the transferred region. Although other explanations cannot strictly be ruled out, the precise and repeatable nature of these changes speaks to intimate, detrimental interactions between deleted loci and the background genome. Along these lines, a recent report demonstrated that independent, massive deletions can occur on a megaplasmid within *Methylobacterium extorquens* during laboratory adaptation and provided strong evidence that the target of selection was not minimization of DNA content [68]. Gene expression and plasmid copy number are also common targets of selection during integration of transferred regions, and might be especially prone to phenotypic convergence hiding underlying genotypic diversity [69–71]. As with deletions described above, although it is straightforward to imagine that the selective target for such changes is energy efficiency, the specific targets of selection are likely complex and difficult to tease apart. Lastly, multiple studies have reported shifts in the conjugative abilities of plasmids after repeated transfer, pointing to an intriguingly direct trade-off between pathways enabling HGT and the costs themselves [16,71,72].

Evolutionary trajectories are heavily influenced by epistatic interactions and order of appearance between new mutations, even when these changes are selectively neutral [9,63]. Costs associated with independent transfer events and the resultant compensatory mutations can therefore funnel populations towards vastly divergent ecological outcomes, especially so when the costs themselves have direct ecological consequences (i.e., ribosomal sequestration [73]).

Outcomes might also depend on population sizes, because even subtle fitness costs could shift the chances of newly acquired regions being lost through genetic drift. Alternatively, the large population sizes necessary for selection to act on the smalles of fitness costs also enable clonal interference between multiple compensatory mutations [74].

Amelioration of costs could affect future evolutionary potential by directly altering regulatory and physiological networks within cells. If recently acquired regions interact or interfere with gene regulation, compensation for fitness costs could involve rewiring chromosomally encoded regulatory circuits, leading to regulatory differences between closely related isolates. Likewise, if foreign DNA sensitizes general stress response pathways, costs associated with HGT could tip the balance to make cells more susceptible to environmental stresses that trigger these pathways, including antibiotics. Such a finding would directly confront the general assumption that HGT typically increases antibiotic resistance and could enable novel or more efficacious treatments. Going forward, compensatory changes to dampen sensitization could end up rendering stress pathways less responsive to environmental challenges. Lastly, costs of HGT can generate phenotypic correlations that would otherwise not exist, setting the stage for instances where selection acting on one phenotype could enable niche expansion in unselected environments. For instance, if HGT lowers overall growth rates as well as thermal tolerance, thermal niche expansion could occur in the absence of selection at higher temperatures simply through compensation of the general fitness costs.

Concluding remarks

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. Although 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 transferred regions are molded to fit within each genome and each genome is in turn altered to accommodate foreign DNA. As understanding of the intricacies of deleterious effects across HGT events increases, these costs can be incorporated into evolutionary models of microbial populations, used to establish further laboratory experiments to validate and quantify hypotheses underlying the costs, and provide a general framework to illuminate the role of HGT across populations.

Acknowledgments

Many thanks go to three anonymous reviewers for their insights and recommendations as well as to Elizabeth Arnold, Corbin Jones, Tory Hendry, and Kevin Hockett for reviewing prior versions of this manuscript. Thank you to Kevin Dougherty for thoughtful discussions through the development of this manuscript. I apologize to all authors whose work was not included within this review owing to space restrictions. D.A.B. was supported through startup funds from the University of Arizona.

References

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 Schonknecht, G. *et al.* (2013) Gene transfer from bacteria and Archaea facilitated evolution of an extremophilic eukaryote. *Science* 339, 1207–1210
4 Moran, N.A. and Jarvik, T. (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* 328, 624–627
5 Yoshida, S. *et al.* (2010) Horizontal gene transfer by the parasitic plant *Striga hermonthica*. *Science* 328, 1128
6 Treangen, T.J. and Rocha, E.P.C. (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 7, e1001284
7 Schaack, S. *et al.* (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* 25, 537–546
8 Park, C. and Zhang, J. (2012) High expression hampers horizontal gene transfer. *Genome Biol. Evol.* 4, 523–532
9 Chou, H.H. *et al.* (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332, 1190–1192
10 Engelberg-Kulka, H. and Glaser, G. (1999) Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu. Rev. Microbiol.* 53, 43–70
11 Diaz-Ricci, J.C. and Hernández, M.E. (2000) Plasmid effects on *Escherichia coli* metabolism. *Crit. Rev. Biotechnol.* 20, 79–108
12 Gonçalves, G.A. *et al.* (2011) Rational engineering of *Escherichia coli* strains for plasmid biopharmaceutical manufacturing. *Biotechnol. J.* 7, 251–261
13 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
14 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
15 Subbiah, M. *et al.* (2011) Selection pressure required for long-term persistence of blaCMY-2-positive IncA/C plasmids. *Appl. Environ. Microbiol.* 77, 4486–4493
16 Harrison, E. and Brockhurst, M.A. (2012) Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends Microbiol.* 20, 262–267
17 Gaillard, M. *et al.* (2008) Host and invader impact of transfer of the *clc* genomic island into *Pseudomonas aeruginosa* PAO1. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7058–7063
18 Sato, T. and Kuramitsu, H. (1998) Plasmid maintenance renders bacteria more susceptible to heat stress. *Microbiol. Immunol.* 42, 467–469
19 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
20 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. U.S.A.* 102, 15971–15976
21 Sorek, R. *et al.* (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318, 1449–1452
22 Creevey, C.J. *et al.* (2011) Universally distributed single-copy genes indicate a constant rate of horizontal transfer. *PLoS ONE* 6, e22099
23 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
24 Pál, C. *et al.* (2005) Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* 37, 1372–1375
25 Green, B. *et al.* (2012) Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mob. DNA* 3, 3
26 Rabinovich, L. *et al.* (2012) Prophage excision activates *Listeria* competence genes that promote phagosomal escape and virulence. *Cell* 150, 792–802
27 Hayes, F. (2003) Transposon based strategies for microbial functional genomics and proteomics. *Annu. Rev. Genet.* 37, 3–29
28 Darling, A.E. *et al.* (2008) Dynamics of genome rearrangement in bacterial populations. *PLoS Genet.* 4, e1000128
29 Matthews, T.D. *et al.* (2010) Chromosomal rearrangements formed by *rrn* recombination do not improve replichore balance in host-specific *Salmonella enterica* serovars. *PLoS ONE* 5, e13503

30 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. U.S.A.* 108, 680–685
31 Jürgen, B. *et al.* (2010) Quality control of inclusion bodies in *Escherichia coli*. *Microb. Cell Fact.* 9, 41
32 Narra, H.P. *et al.* (2008) Structural features and the persistence of acquired proteins. *Proteomics* 8, 4772–4781
33 Drummond, D.A. and Wilke, C.O. (2009) The evolutionary consequences of erroneous protein synthesis. *Nat. Rev. Genet.* 10, 715–724
34 Kohanski, M.A. *et al.* (2008) Mistranslation of membrane proteins and two-component system activation trigger antibiotic-mediated cell death. *Cell* 135, 679–690
35 Bragg, J.G. and Wagner, A. (2009) Protein material costs: single atoms can make an evolutionary difference. *Trends Genet.* 25, 5–8
36 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
37 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
38 Pal, C. *et al.* (2005) Horizontal gene transfer depends on gene content of the host. *Bioinformatics* 21, ii222–ii223
39 Vavouri, T. *et al.* (2009) Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* 138, 198–208
40 Yang, J.R. *et al.* (2012) Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc. Natl. Acad. Sci. U.S.A.* 109, E831–E840
41 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
42 Bonomo, J. and Gill, R.T. (2005) Amino acid content of recombinant proteins influences the metabolic burden response. *Biotechnol. Bioeng.* 90, 116–126
43 Chen, Y. *et al.* (2005) Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. *PLoS Biol.* 3, e229
44 Shintani, M. *et al.* (2009) Response of the *Pseudomonas* host chromosomal transcriptome to carriage of the IncP-7 plasmid pCAR1. *Environ. Microbiol.* 12, 1413–1426
45 Schalk, I.J. *et al.* (2011) New roles for bacterial siderophores in metal transport and tolerance. *Environ. Microbiol.* 13, 2844–2854
46 Azad, R.K. and Lawrence, J.G. (2011) Towards more robust methods of alien gene detection. *Nucleic Acids Res.* 39, e56
47 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. U.S.A.* 99, 3695–3700
48 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. U.S.A.* 5, e1000513
49 Raghavan, R. *et al.* (2012) A selective force favoring increased G+C content in bacterial genes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14504–14507
50 Gingold, H. and Pilpel, Y. (2011) Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* 7, 1–13
51 Agashe, D. *et al.* (2013) Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *Mol. Biol. Evol.* 30, 549–560
52 Tuller, T. *et al.* (2011) Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Res.* 39, 4743–4755
53 Cordero, O.X. and Hogeweg, P. (2009) The impact of long-distance horizontal gene transfer on prokaryotic genome size. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21748–21753
54 Humphrey, B. *et al.* (2012) Fitness of *Escherichia coli* strains carrying expressed and partially silent IncN and IncP1 plasmids. *BMC Microbiol.* 12, 53
55 Singer, A. *et al.* (2009) DNA plasmid production in different host strains of *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* 36, 521–530
56 Silva, R.F. *et al.* (2011) Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. *PLoS Genet.* 7, e1002181
57 Miyakoshi, M. *et al.* (2009) High-resolution mapping of plasmid transcriptomes in different host bacteria. *BMC Genomics* 10, 12
58 Slater, F.R. *et al.* (2008) Progress towards understanding the fate of plasmids in bacterial communities. *FEMS Microbiol. Ecol.* 66, 3–13
59 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
60 Platt, T.G. *et al.* (2012) A cooperative virulence plasmid imposes a high fitness cost under conditions that induce pathogenesis. *Proc. Biol. Sci.* 279, 1691–1699
61 Grote, J. *et al.* (2012) Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* 3, e00252-12
62 Grzymski, J.J. and Dussaq, A.M. (2011) The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J.* 6, 71–80
63 Woods, R.J. *et al.* (2011) Second-order selection for evolvability in a large *Escherichia coli* population. *Science* 331, 1433–1436
64 Tenaillon, O. *et al.* (2012) The molecular diversity of adaptive convergence. *Science* 335, 457–461
65 Lawrence, J.G. and Ochman, H. (1997) Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* 44, 383–397
66 Lercher, M.J. and Pal, C. (2008) Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol. Biol. Evol.* 25, 559–567
67 Amoros-Moya, D. *et al.* (2010) Evolution in regulatory regions rapidly compensates the cost of nonoptimal codon usage. *Mol. Biol. Evol.* 27, 2141–2151
68 Lee, M-C. and Marx, C.J. (2012) Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genet.* 8, e1002651
69 Chou, H-H. and Marx, C.J. (2012) Optimization of gene expression through divergent mutational paths. *Cell Rep.* 1, 133–140
70 Sota, M. *et al.* (2010) Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein. *ISME J.* 4, 1568–1580
71 Dahlberg, C. and Chao, L. (2003) Amelioration of the cost of conjugative plasmid carriage in *Escherichia coli* K12. *Genetics* 165, 1641–1649
72 Turner, P.E. *et al.* (1998) Tradeoff between horizontal and vertical modes of transmission in bacterial plasmids. *Evolution* 52, 315–329
73 Dethlefsen, L. and Schmidt, T.M. (2007) Performance of the translational apparatus varies with the ecological strategies of bacteria. *J. Bacteriol.* 189, 3237–3245
74 Lee, M-C. and Marx, C.J. (2013) Synchronous waves of failed soft sweeps in the laboratory: remarkably rampant clonal interference of alleles at a single locus. *Genetics* 193, 943–952
75 Drummond, D.A. and Wilke, C.O. (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134, 341–352
76 Doyle, M. *et al.* (2007) An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science* 315, 251–252
77 Skennerton, C.T. *et al.* (2011) Phage encoded H-NS: a potential achilles heel in the bacterial defence system. *PLoS ONE* 6, e20095
78 Petrova, V. *et al.* (2009) An SOS inhibitor that binds to free RecA protein: the PsiB protein. *Mol. Cell* 36, 121–130
79 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
80 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
81 Otto, S.P. (2009) The evolutionary enigma of sex. *Am. Nat.* 174 (Suppl. 1), S1–S14
82 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
83 Hatfull, G.F. (2008) Bacteriophage genomics. *Curr. Opin. Micribol.* 11, 447–453
84 Liu, J. *et al.* (2006) Protein repertoire of double-stranded DNA bacteriophages. *Virus Res.* 117, 68–80
85 Harrison, P.W. *et al.* (2010) Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol.* 18, 141–148
86 Smillie, C. *et al.* (2010) Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* 74, 434–452
87 Baltrus, D.A. *et al.* (2011) Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathog.* 7, e1002132