Review

Correspondence

Rob Knight

Rob.Knight@colorado.edu

Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns

Jesse R. Zaneveld,¹ Diana R. Nemergut² and Rob Knight³¹Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309, USA²Institute of Arctic and Alpine Research (INSTAAR) and Environmental Studies Program, University of Colorado, Boulder, CO 80309, USA³Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA

Detecting patterns of horizontal gene transfer (HGT) in genomic sequences is an important problem, with implications for evolution, ecology, biotechnology and medicine. Extensive genetic, biochemical and genomic studies have provided a good understanding of sequence features that are associated with many (though not all) known mobile elements and mechanisms of gene transfer. This information, however, is not currently incorporated into automated methods for gene transfer detection in genomic data. In this review, we argue that automated annotation of sequence features associated with gene transfer mechanisms could be used both to build more sensitive, mechanism-specific compositional models for the detection of some types of HGT in genomic data, and to ask new questions about the classes of genes most frequently transferred by each mechanism. We then summarize the genes and sequence features associated with different mechanisms of horizontal transfer, emphasizing those that are most useful for distinguishing types of transfer when examining genomic data, and noting those classes of transfers that cannot be distinguished in genomic data using existing techniques. Finally, we describe software, databases and algorithms for identifying particular classes of mobile elements, and outline prospects for better detection of HGT based on specific mechanisms of transfer.

Introduction

The capacity of microbes to exchange genes through horizontal gene transfer (HGT) is an important phenomenon with implications for ecology, evolution, biotechnology and medicine (Ochman *et al.*, 2000; Frost *et al.*, 2005). HGT can promote adaptation to new environments by allowing molecular innovations to be traded within ecological niches. For example, HGT has enhanced bacterial adaptation to life in environments containing heavy metals (Liebert *et al.*, 1999) and toxic organic compounds (Top & Springael, 2003). Unfortunately, HGT has also allowed adaptation to the human host environment, both by spreading pathogenicity determinants (Canchaya *et al.*, 2003) and by allowing pathogenic bacteria to acquire resistance to multiple classes of antibiotics (Walsh, 2006). Thus, HGT has diverse and important societal implications, both positive and negative, and deserves detailed study.

An understanding of how microbial genomes have been shaped by HGT events has been pursued in parallel using genetic, biochemical and computational approaches. Genetic and biochemical approaches to HGT have succeeded in elucidating many of the mechanisms by which HGT is accomplished (Curcio & Derbyshire, 2003; Averhoff, 2004; Chen *et al.*, 2005). However, because the pattern of horizontal transfers that has affected each genome is unique, and because these horizontal transfers have occurred over very long timescales, the history of HGT in a particular genome cannot be uncovered in an experimental setting. Instead, past horizontal transfer events must be inferred from the molecular fingerprints that those events have left on extant microbial genomes. These inferences are made using computational approaches.

Inferring HGT from genomic data

The detection of HGT in genomic data is currently accomplished using two main approaches: phylogenetic and compositional. Phylogenetic methods attempt to examine the distribution of, or relationships between,

Abbreviations: BLAST, basic local alignment search tool; DUS, DNA uptake sequence; GTA, gene transfer agent; HGT, horizontal gene transfer; ICE, integrative and conjugative element; IS, insertion sequence; SSV1, *Sulfolobus* spindle virus 1; TIIS, type II secretory system; TIVSS, type IV secretory system; USS, uptake signal sequence.

genes in multiple taxa. For example, discordance between the phylogeny of a gene and the species phylogeny may indicate horizontal transfer. Such results should be treated with caution, however, since gene duplication and differential loss or selection can generate similar patterns (Syvanen, 1994; Page & Charleston, 1997). Additionally, even in a genome in which all genes share the same history, the sequence of some genes may not support the true phylogeny. This effect is especially important where the phylogenetic signal is weak, and thus susceptible to being overwhelmed by random noise.

Techniques related to these phylogenetic methods include distributions of BLAST hits (e.g. Nelson *et al.*, 1999; Lander *et al.*, 2001), inference of gene presence/absence within species phylogenies (e.g. Jordan *et al.*, 2001; Mirkin *et al.*, 2003), and ratios of evolutionary distances (e.g. Farahi *et al.*, 2003; Kechris *et al.*, 2006). These techniques, however, have been shown to detect only a subset of horizontally transferred genes, and therefore should be treated as approximations to, not replacements for, phylogenetic methods (Ragan, 2001; Kinsella & McNerney, 2003; Ragan *et al.*, 2006).

Compositional methods, in contrast, examine sequence characteristics that vary in different taxa but are relatively conserved within a taxon. If a gene or genomic fragment possesses unusual sequence characteristics, it may have been transferred from a taxon in which those characteristics are typical. Unlike phylogenetic methods, compositional methods do not require alignment of homologous sequences, and are therefore better suited to examine poorly conserved or very rapidly evolving genes. Sequence characteristics that have been used to study HGT include GC content (in the whole gene or at the third position in each codon), the codon adaptation index, amino acid usage, relative synonymous codon usage and dinucleotide usage (Sharp & Li, 1987; Lawrence & Ochman, 1997, 1998; Karlin *et al.*, 1998; Hooper & Berg, 2002). Markov models, including frame-dependent Markov models (Nakamura *et al.*, 2004) and variable-order Markov models (Vernikos & Parkhill, 2006) have also been used. One disadvantage of compositional methods is that transfer between related or unrelated strains with similar compositional characteristics will not be detected. Conversely, if a sequence characteristic varies within a genome, then classes of genes, such as those encoding ribosomal proteins, that have unusual compositions may appear to have been horizontally transferred. Even when the composition of a transferred gene is initially distinct from the composition of the recipient genome, the composition of the gene will drift toward that of the recipient genome over time. Thus, the traces of ancient transfers may be obliterated in sequences that have had time to equilibrate fully (Lawrence & Ochman, 1997).

Benefits of differentiating between HGT mechanisms

Current phylogenetic and sequence composition-based methods for detecting horizontally transferred genes in

genomic data treat all horizontally transferred genes as equivalent. We argue that these computational approaches to identifying HGT could be improved by incorporating recent insights from genetic and biochemical studies into the mechanisms that allow gene transfers to occur. Molecular biology has uncovered sequence features associated with many, though by no means all, HGT mechanisms, and these sequence features can be exploited to categorize many putatively transferred genes according to the mechanism of transfer.

Categorizing HGT events that can be identified by mechanism could provide three important benefits. First, such categorization could allow testing of whether the genes transferred by different mechanisms vary in properties such as composition or cellular function. Second, automatic identification of sequence features could allow for the construction of high-confidence datasets of horizontally transferred genes. Third, such categorization could allow the construction of mechanism-specific models that may detect HGT with greater sensitivity than the general techniques currently employed.

The first benefit of categorizing HGT genes by transfer mechanism is that such categorization could allow for quantification of differences in the properties of mechanisms that can be identified by sequence features. Different mobile elements or mechanisms of transfer may tend to transmit distinct classes of genes, predominate in specific environmental conditions, or transfer genes between hosts separated by variable phylogenetic distances. It is further possible, as suggested by one of our reviewers, that genes with different degrees of divergence in sequence or in compositional characteristics may be transferred by specific mechanisms (although the latter case may be difficult to disentangle from compositional differences induced by residency in a particular mobile element). It is also possible that several of these factors may interact in complex ways to determine which genes are transferred by a particular mechanism. Exploring these potential differences between mechanisms of transfer may help us to understand the different evolutionary implications of distinct classes of transfer events.

The second benefit of categorizing putative HGT genes by their mechanism of transfer is that such categorization may be used to generate a higher-confidence pool of HGT genes from a large set of putative HGT genes detected using phylogenetic or compositional methods. This can be accomplished by searching the region around each putatively transferred gene for the presence or absence of sequence features that indicate a mechanism of HGT. For example, if a gene is found to be integrated into a tRNA gene, and several phage genes are located nearby, then not only has very strong additional support been generated for the hypothesis of transfer, but a specific mechanism (phage transduction) can be proposed. However, the converse does not hold. Because not all mechanisms of transfer can be distinguished on the basis of sequence features alone,

the absence of sequence features indicating a known mechanism does not by itself demonstrate that a gene was not transferred (Table 1). The pool of high-confidence HGT genes generated in this manner will necessarily be limited to genes transferred by mechanisms identifiable from their sequence features.

Although these approaches are not new, confirmation of HGT predictions in this manner has mostly been done manually, in an ad hoc fashion. Therefore, it could only be practically applied to small sets of genes that are of particular interest. To the best of our knowledge, no major compositional or phylogenetic methods use information about mobile elements in a systematic way to confirm predictions of HGT. One step in this direction was taken by William Hsiao and coworkers, who developed the IslandPath software (Hsiao *et al.*, 2003). IslandPath performs several kinds of compositional analyses, and annotates integrases and transposases in a genome. Information about the location of genes associated with mobility was later used to provide support for genomic islands predicted on the basis of unusual dinucleotide usage (Hsiao *et al.*, 2005). This approach could also be used to estimate what proportion of putative gene transfers are accounted for by mechanisms identifiable from their sequence features. Genes with evidence of transfer but that lack sequence features associated with a known transfer mechanism could be transfers effected by known mechanisms that are not identifiable by sequence features, false positives of the compositional or phylogenetic method used to detect transfer, or genes transferred by a previously unknown mechanism.

The third benefit of categorizing HGT genes by transfer mechanism is that distinguishing between different types of transfer could allow for the construction of specialized, mechanism-specific compositional methods for detecting HGT events effected by different mechanisms. Existing compositional methods are general, and attempt to detect all HGT events. Categorizing HGT events by mechanism would allow for mechanism-specific compositional models to be built, which could improve sensitivity when searching for genes transferred by that mechanism. For example, the GC content of plasmids tends to be lower than that of the recipient genome; thus we might expect that building

compositional models for this element would improve the detection of genes transferred by plasmid integration (van Passel *et al.*, 2006).

Thus, we propose that automated annotation of the mechanism of gene transfer using sequence features present in the recipient genome will, where it is possible, provide a useful complement to existing compositional and phylogenetic methods for studying HGT.

Mobile DNA elements are mosaics of functional modules

One significant barrier to incorporating information about the mechanism of HGT into genomic analyses is the sheer diversity of mobile elements that are known to exist. Mobile elements as varied as conjugative transposons, transposable prophages, integrative plasmids and mobile integrons appear on first inspection to be extremely difficult to detect and categorize systematically. Indeed, early attempts to classify mobile elements as phage, plasmid or transposon were thwarted by the discovery of new elements, such as conjugative transposons, which possess unique combinations of previously known mobility determinants (Franke & Clewell, 1981; Osborn & Boltner, 2002). A key insight that has helped to organize and interpret this diversity of gene transfer mechanisms is that mobile genomic elements can be thought of as the sum of a set of relatively independent and exchangeable functional units (Osborn & Boltner, 2002; Toussaint & Merlin, 2002). These functional units can broadly be categorized as promoting intercellular or intracellular mobility, or replication, selection, stability and maintenance (Toussaint & Merlin, 2002).

Rather than creating an ever-expanding list of new categories of mobile elements, the mechanism-centric approach to mobile element diversity attempts to link particular mobility functions to particular sets of genes and sequence features. Thus, although the combinatorial nature of functional modules composing mobile elements makes classification difficult at the level of the element as a whole, this approach could allow for the functional capabilities of a particular mobile element to be predicted from its sequence, and provide the necessary framework to identify

Table 1. Proposed interpretation of combined phylogenetic/compositional data and sequence features

Phylogenetic or compositional evidence of transfer	Mechanism-specific sequence features	Interpretation
+	+	Strong support for hypothesis of transfer
+	–	False positive, transfer by a known mechanism lacking diagnostic sequence features, or transfer by a novel transfer mechanism
–	+	False negative, or potentially mobile sequence that has not been transferred
–	–	Strong evidence that no recent, long-range transfers occurred; transfers over short phylogenetic distances by mechanisms without diagnostic sequence features remain a possibility

and describe mobile elements that do not fit into classical categories. This way of thinking about the diversity of mobile elements is now the basis of the ACLAME database, which seeks to annotate mobile genetic elements on the basis of the functional modules they contain (Leplae *et al.*, 2004).

In order to facilitate comparisons of the types of genes transferred by different mechanisms, we review the known mechanisms of inter- and intragenomic gene transfer, emphasizing the distinct evolutionary pressures and mechanistic restraints each imposes, and the characteristics likely to prove most useful in distinguishing HGT mechanisms when examining genomic data.

Detecting mechanisms of intercellular DNA mobility using genomic data

In order for genes to be transferred between microbial taxa, genomic information must be transferred between microbial cells. In most cases, this transfer of information is

accomplished by the exchange of DNA (although some phage can transduce genomic information via an RNA intermediate). Intercellular DNA exchange can be facilitated by functional elements encoded in the donor genome, in the recipient genome, on mobile elements, or some combination of the three (Fig. 1). Mechanisms of intercellular DNA transfer include conjugation, phage transduction, and transformation.

Conjugation

Conjugation involves the transfer of DNA between cells in a contact-dependent fashion. Plasmids, conjugative transposons (such as Tn916), regions of bacterial chromosomes, and integrative and conjugative elements (ICEs) all transfer in this fashion (Tatum & Lederberg, 1947; Franke & Clewell, 1981; Burrus *et al.*, 2002). Conjugative DNA transfer is widespread, having been detected in both Gram-positive and -negative bacteria, and in Crenarchaeote but not Euryarchaeote archaea (Prangishvili *et al.*, 1998 and references therein; Stedman *et al.*, 2000).

Fig. 1. Inter- and intracellular mechanisms of gene transfer. Processes are shown in italics. Notably, several types of mobile element may move by more than one mechanism of gene transfer. Conversely, particular mechanisms of gene transfer can mobilize multiple types of element. (Many known classes of mobile elements, such as conjugative transposons, are omitted for clarity.)

DNA transfer during conjugation is effected by type IV secretory systems (TIVSSs). TIVSSs transfer only single-stranded DNA, so the DNA to be transferred must be separated from its complementary strand prior to transfer. In order to separate the two strands, the DNA is nicked. This nicking is achieved by a relaxase, which binds to a *cis* element called an origin of transfer (*oriT*). Relaxase binding promotes the formation of a relaxosome complex, and nicks the DNA at a conserved *nic* motif. Although relaxases can be supplied *in trans* (for example by a helper plasmid), conjugative DNA elements must contain an origin of transfer *in cis* in order to be moved through the conjugation apparatus (Grohmann *et al.*, 2003 and references therein). Because *oriT* elements appear to be an absolute *cis* requirement for conjugal transfer, these sequences are good candidates for the identification of mobile elements or genomic regions that may be capable of movement by conjugative transfer (Table 2). A second property that makes these sequences good candidates for genomic markers of transfer by conjugation is their conservation (Fig. 2). The *nic* motifs of *oriT* sequences necessary for conjugative transfer fall into five conserved classes. Origins of transfer also fall into five conserved families, and alignments of these sequences are available (e.g. Lanka & Wilkins, 1995; Zechner *et al.*, 2000). Transfer by conjugation could be detected by searching the region around each putatively transferred gene for origins of transfer using similarity searches against profiles of sequences from each *oriT* family (Altschul *et al.*, 1997).

A conjugation-like DNA transfer mechanism in *Streptomyces*. A substantial body of information has accumulated to suggest that a novel conjugation-like mechanism is at work in the plasmids of *Streptomyces*. In the *Streptomyces* plasmids pIJ101 and pJV1, a *cis*-acting locus of transfer (*clt*) is required for plasmid mobilization (Servin-Gonzalez, 1996 and references therein). These *clt* sequences lack sequence similarity to origin of transfer (*oriT*) sequences required for traditional conjugation, and appear to act by a novel mechanism. Unlike conventional conjugation, which requires a number of components, only a single *tra* gene is necessary *in trans* for conjugation to occur (Grohmann *et al.*, 2003) (Table 2). The *tra* genes from several *Streptomyces* plasmids appear to be related to the SpoIIIE/FtsK septal DNA transport proteins. SpoIIIE is involved in the transport of double-stranded DNA into the pre-spore of *Bacillus subtilis* (Errington, 2001), suggesting that TraA may also transfer double-stranded DNA (dsDNA) rather than single-stranded DNA as in traditional conjugation. Experiments using restriction enzymes support this notion. Conjugative transfer by TraA is sensitive to dsDNA-specific *SalI* restriction digestion, and is thus also likely to involve the transfer of dsDNA (Possoz *et al.*, 2001). So far, the novel conjugation mechanism of the pIJ101 and pJV1 plasmids of *Streptomyces* has not been explored thoroughly enough to provide sequences that could be exploited in genomic studies. Indeed, although direct and/or inverted repeats appear to play a role in all known examples, the sequence

Table 2. Genes and sequence features necessary for transfer by each intercellular transfer mechanism

oriT, origin of transfer; *mob*, relaxase gene; *tra*, transfer genes involved in conjugative transfer, which may include relaxases, TIVSS homologues, mating pair formation proteins, coupling proteins, etc. (Grohmann *et al.*, 2003). *attB*, host phage attachment site; *attP*, phage attachment site; *clt*, chromosomal locus of transfer. *cis* factors are those required *in cis* by transferred DNA, *trans* factors are those that are required *in trans* by the transferred DNA. Mobile elements frequently encode genes for functions not listed here. For example, many conjugative plasmids encode genes for plasmid partitioning, origins of replication, etc., in addition to features allowing conjugative transfer.

Mechanism	Donor genome		Recipient genome	
	<i>cis</i> factors	<i>trans</i> factors	<i>cis</i> factors	<i>trans</i> factors*
Specialized transduction	<i>attB</i> , integrase, phage structural genes	—	<i>attP</i>	Cell surface receptor
Generalized transduction	—	Integrase, phage structural genes†	—	Cell surface receptor
Gene transfer agents	—	Phage structural genes†	—	RecA, cell surface receptor?
Conjugation	<i>oriT</i>	<i>tra</i> , <i>mob</i> †	—	—
Conjugation-like transfer in <i>Streptomyces</i>	<i>clt</i>	<i>tra</i>	—	—
Transformation	—‡	—	—	TISS, RecA

*Although integration host factor (IHF) facilitates λ phage integration, this factor is not listed because it is not clear that IHF homologues are always required for integrative site-specific recombination (Cho *et al.*, 1999).

†Although phage structural genes are required *in trans* for transduction to occur, these genes will not be transferred by gene transfer agents, and will frequently also not be transferred by generalized transduction. Similarly, although a relaxase is required *in trans* for conjugative transfer to occur, it need not be encoded on the mobile element itself, and thus may not be transferred to the recipient genome.

‡For uptake by *Neisseria* or *Helicobacter* a DNA uptake sequence (DUS) is required.

Fig. 2. Sequence features associated with mechanisms of inter- and intracellular gene transfer. In the diagram of origin of transfer *nic* motifs, an asterisk indicates a cutting site. In the diagram of integron integrases, the shaded regions are unique additional domains. Prophage λ layout adapted from Court *et al.* (2007). Sequences, consensus sequences and alignments for *oriT* *nic* motifs, and IntI additional domains are from Lanka & Wilkins (1995) and Messier & Roy (2001) respectively.

and secondary structure of *clt* sequences appear to be highly variable (Servin-Gonzalez, 1996; Franco *et al.*, 2003). Developing methods for detecting and annotating *clt* sites *in silico* remains a challenge for the future.

Phage transduction

Phage transduction as a mechanism of genetic exchange was discovered by Zinder & Lederberg (1952) in *Salmonella typhimurium* (Table 3). Phage gain entry to their microbial host cell by binding to extracellular receptors on the cell surface. Following entry into the host cell, lytic phage typically replicate numerous copies of their genome, and then exit the cell en masse, causing cell lysis. In addition to immediate replication and destruction of the host cell, temperate or lysogenic phage are also capable of integrating into the host genome, where they can lie dormant, replicating along with the host.

When triggered (e.g. by DNA damage), these lysogenic prophage can excise themselves from the host genome, rapidly replicate, and exit the cell by lysis. Occasionally, however, errors in the excision of prophage from the host genome can result in the packaging of host DNA along

with or instead of the phage genome. When the phage containing microbial DNA inserts itself into a new host, the DNA from the previous host can be transferred to a new genome. Although traditionally it was expected that phage capable of infecting across species should be rare, broad host-range phage (including generalized transducing phage) capable of infecting hosts across genera, or even classes, have been detected (Jensen *et al.*, 1998).

Phage transduction can be broadly separated into specialized and generalized transduction (Table 2).

Specialized transduction. Specialized transduction occurs in phage that preferentially integrate at particular attachment sites in the microbial host genome (*attB*) through integrase-mediated recombination with a second attachment site in the phage genome (*attP*). If such a phage is incorrectly excised from its host genome, then small regions of host DNA on either side of the prophage may be packaged along with the prophage genome (Canchaya *et al.*, 2003). Since the site of phage integration is highly specific as long as the host attachment site is present, the variety of genes that can be transferred from a particular host by specialized transduction is usually quite limited.

Table 3. Percentage of gene transfer and mobile elements reported to be present in selected genomes

The number of genes in each genome and the estimate of the percentage of transfer are from the analysis of Nakamura *et al.* (2004). It should be noted, however, that other large-scale surveys of gene transfer reach different results, depending on the methodology used (Ge *et al.*, 2005; Hsiao *et al.*, 2005; Choi & Kim, 2007). Numbers of transposases and integrases from IslandPath were manually counted from the display on the updated IslandPath website (<http://www.pathogenomics.sfu.ca/islandpath/>) and so should be considered approximate (Hsiao *et al.*, 2003). Numbers of ISs are from the ISFinder website (<http://www-is.biotoul.fr/>) (Siguier *et al.*, 2006).

	<i>Escherichia coli</i> K-12 MG1655	<i>Salmonella</i> <i>typhimurium</i> LT2	<i>Helicobacter</i> <i>pylori</i> J99	<i>Deinococcus</i> <i>radiodurans</i> R1*	<i>Vibrio cholerae</i> serotype 01 N16961*	<i>Sulfolobus</i> <i>solfatarius</i> P2
No. of genes in genome	4278	4440	1487	2937	3,790	2977
% Transfer (Nakamura <i>et al.</i> , 2004)	16.90 %	16.50 %	8.10 %	14.40 %	16.70 %	8.70 %
Natural transformation reported?	—	—	+ (Hofreuter <i>et al.</i> , 2001)	+ (Tigari & Moseley, 1980)	+ (Meibom <i>et al.</i> , 2005)	—
Transposases (IslandPath)	37	16	2	33	22	184
ISs (ISFinder)	42	19 (10 partial)	10 (9 partial)	37 (10 partial)	16	191
Integrases (IslandPath)	7	8	3	2	5	7
Integrases near tRNA (IslandPath)	2	3	0	1	2	6
Minimal no. of prophage (Casjens, 2003)	11	7	2	0	1	NA
Reported integrons	0†	0†	0	0	1‡	0

NA, Not applicable.

*Numbers of mobile elements and genes reflect the total on both chromosomes.

†Although no integrons are present in the sequenced genome, they have been found in related strains (van Essen-Zandbergen *et al.*, 2007).

‡Integron reported by Heidelberg *et al.* (2000).

Nonetheless, pathogenicity determinants have been shown to have been acquired in this manner (Canchaya *et al.*, 2003 and references therein).

Generalized transduction. In contrast to specialized transduction, generalized transduction can in principle facilitate the movement of any gene. The best-studied example of generalized transduction is Mu phage. Mu phage integrates randomly, and upon excision packages its DNA by recognition of a *pac* site within the Mu prophage sequence (Harel *et al.*, 1990). Occasionally, however, Mu may package a segment of DNA from its bacterial host instead. In that case the Mu integrase is not packaged, and the transduced DNA must enter the recipient genome by homologous recombination in a RecA-dependent fashion (Bushman, 2002). Even when acting normally, Mu can transduce smaller regions of host DNA. The Mu head can hold more DNA than the Mu phage itself encodes; up to 100 bp of DNA upstream of the prophage, and 2000 bp downstream of the prophage may be included to fill the phage head (Bushman, 2002). Following infection of a new

bacterial host, this extra genetic material can be integrated alongside the phage genome.

Genes transduced by phage may be detected by their proximity to genes with sequence similarity to known phage integrases and structural genes, the presence of direct repeats, and integration near tRNA genes (Williams, 2002) (Table 4). Two programs for detecting prophage in genomic data include Phage_Finder and ProphageFinder (Bose & Barber, 2006; Fouts, 2006). Phage_Finder functions by using BLAST and phage-specific sequences to search genomes for regions containing phage genes. It then removes putative integrons and bacteriocins, and is also capable of further subclassifying phage (for example, by identifying Mu-like phage) (Fouts, 2006). ProphageFinder uses BLASTX to search user-defined regions for phage genes (Bose & Barber, 2006).

Gene transfer agents. Interestingly, both archaea and bacteria contain defective prophages that have lost the ability to package phage genes, and instead package and

Table 4. Selected sequence features associated with HGT mechanisms

Mechanism	Feature	Available programs	Citations
Transduction	Phage structural genes (integrases, head proteins)	Phage_Finder, ProphageFinder	Bose & Barber (2006); Fouts (2006)
	Location in genome (near tRNA or tmRNA)	IslandPath	Williams (2002); Hsiao <i>et al.</i> (2003)
Conjugation	Origin of transfer (<i>oriT</i>)		Lanka & Wilkins (1995); Zechner <i>et al.</i> (2000)
	Chromosomal locus of transfer (<i>clt</i>)		Servin-Gonzalez (1996); Franco <i>et al.</i> (2003) and references therein
Transformation	DNA uptake sequence (DUS) (in species that require them)		Davidsen <i>et al.</i> (2004); Ambur <i>et al.</i> (2007)
Integron integration	Integron integrase (<i>intI</i>)	IslandPath, BLAST	Messier & Roy (2001) and references therein
Integron gene cassette	<i>attC</i>	XXR	Rowe-Magnus <i>et al.</i> (2003)
Transposition	Transposase	ISFinder, IslandPath	Hsiao <i>et al.</i> (2003); Siguier <i>et al.</i> (2006)

transduce small random regions of host DNA (Marrs, 1974; Bertani, 1999). These elements are termed gene transfer agents (GTAs), and were originally identified in the photosynthetic purple nonsulphur bacterium *Rhodobacter capsulatus* (Marrs, 1974). Subsequently, GTAs have been identified in the Euryarchaeote *Methanococcus voltae* PS, and in a variety of diverse bacteria (Humphrey *et al.*, 1997; Bertani, 1999; Lang & Beatty, 2007 and references therein). Since no phage integrase is packaged, these small transduced fragments must enter the genome through recombination. At least one phylogenetic analysis suggests that GTAs arose only once in the α -proteobacteria (Lang & Beatty, 2007), although the origin of the archaeal GTA in *Methanococcus voltae* PS almost certainly represents a separate event.

Because GTAs do not transfer phage structural genes or integrases, and because they are believed to rely on homologous recombination to enter the target genome, GTA-mediated transfers are likely to be much more difficult to detect, and to distinguish from other gene transfer mechanisms than traditional phage transduction. Two criteria may nonetheless help to do so. First, regions of DNA packaged into GTAs are believed to be short (≈ 4.5 kb). Secondly, the organism donating DNA by this mechanism must contain GTA genes. Thus, in cases where the donor lineage can be discerned, the possibility of transfer by a GTA could be established by demonstrating that GTAs are common in the donor lineage. For example, the widespread distribution of GTAs in α -proteobacteria suggests that GTA-mediated transfer should be considered a possibility when examining short transferred fragments of α -proteobacterial origin. In the future, identification of the cell-surface receptors for known GTAs could help to identify lineages that have the capability of acquiring DNA in this manner.

Transformation

Transformation is the uptake of circular or linear DNA from the environment (Dubnau, 1999). Natural transformation was first documented in bacteria, a discovery

that led to the finding that DNA is the genetic material (Griffith, 1928; Avery *et al.*, 1944). More recently, natural transformation has also been reported within the archaea *Methanococcus voltae* PS (Bertani & Baresi, 1987), *Methanobacterium thermoautotrophicum* Marburg (Worrell *et al.*, 1988) and *Thermococcus kodakaraensis* KOD1 (Sato *et al.*, 2003), although the transformation mechanism used by the archaea is not yet known.

Transformation has several stages: induction of competence (in strains that are not constitutively competent), availability of DNA in the medium, DNA binding, DNA fragmentation, transport across the outer membrane (in Gram-negative bacteria) or the cell wall (in Gram-positives), conversion to single-stranded DNA, and transport across the cytoplasmic membrane (Dubnau, 1999; Averhoff, 2004; Chen *et al.*, 2005). If the transformed DNA is to be stably maintained in the cell, it must then recombine with host, plasmid or phage DNA or, in the case that a plasmid was taken up from the environment through the transformation apparatus, recircularize (for several excellent reviews, see Lorenz & Wackernagel, 1994; Averhoff, 2004; Chen *et al.*, 2005; Thomas & Nielsen, 2005).

In all naturally transformable bacteria studied to date, with the notable exceptions of *Helicobacter pylori* and *Campylobacter jejuni* (see below), transformation depends on some protein components also utilized in type IV pilus formation, twitching motility, and the type II secretory system (TISS) (Dubnau, 1999; Averhoff, 2004). These proteins are exemplified by the Com proteins of *Bacillus subtilis*. Although components of type IV pili are necessary for natural competence, it appears that the pili themselves are dispensable (Averhoff, 2004). This could be explained if the transformation apparatus utilizes a pseudopilus, which competes for common components with type IV pili: Chen *et al.* (2005) have recently proposed a model for transformation in *B. subtilis* in which the uptake of DNA is driven by the force of repeated cycles of polymerization and depolymerization of the pseudopilus, where depolymerization of the pseudopilus is driven by the

proton-motive force. Once single-stranded DNA has begun to enter the cytoplasm, the binding of single-strand binding (SSB) proteins is proposed to aid uptake through a Brownian ratchet mechanism (Chen *et al.*, 2005 and references therein).

In *Neisseria* and the *Pasteurellaceae*, the transformation apparatus will only take up environmental DNA that contains a specific 9–11 nucleotide sequence. These sequences, called uptake signal sequences (USSs) or DNA uptake sequences (DUSs), are the most common sequences of their length in their genome (Davidsen *et al.*, 2004 and references therein) (Fig. 2). No receptor that recognizes USSs has yet been found. Other naturally transformable bacteria such as *Bacillus subtilis* and *Streptococcus pneumoniae* do not appear to require such sequences (Dubnau, 1999). Likewise no requirement for signal sequences has yet been found in naturally transformable archaea.

The genomes of organisms that require signal sequences to achieve DNA uptake provide useful model systems for studying transformation-mediated HGT, because genes within them lacking a DUS are unlikely to have been acquired by transformation in the recent past (Table 2). Conversely, overrepresentation of DUSs within a class of genes may indicate that those genes are frequently horizontally transferred by transformation. As an example of this type of reasoning, Davidsen *et al.* (2004) demonstrated that DNA uptake signals are overrepresented in repair and recombination genes that promote genome maintenance, suggesting that transformation and recombination may help to protect DNA repair genes from genotoxic stress. This overrepresentation could not be accounted for by the dinucleotide composition of genome maintenance genes, and was found in all genomes known to contain DUSs. No overrepresented DNA sequences of similar length were found in *E. coli*, which is not believed to be naturally transformable (Lorenz & Wackernagel, 1994; Davidsen *et al.*, 2004).

Transformation in *H. pylori* and *C. jejuni*. An unusual system for transformation is found in *Helicobacter pylori* (Table 3). Transformation in *Helicobacter* is mediated by a TIVSS that contains components homologous to many elements of the *Agrobacterium tumefaciens* conjugal T-DNA transfer system (Hofreuter *et al.*, 2001). A transmissible TIVSS involved in transformation has also been reported on a plasmid of *Campylobacter jejuni* (Bacon *et al.*, 2000). As yet, no special signal sequences have been found that are required for DNA uptake through the TIVSS of *Helicobacter* or *Campylobacter*.

Detecting mechanisms of intracellular DNA and RNA mobility in genomic data

The transfer of DNA between cells is only one aspect of the horizontal transfer process; movement of DNA and RNA within cells is also important in the generation of variation in bacteria. Intracellular DNA and RNA mobility is crucial

for allowing genomic mobile elements to escape the donor genome prior to intercellular transfer, and to entering a recipient genome after transfer. Intracellular DNA mobility can also transfer DNA between the microbial genome and a plasmid or phage genome, and can promote genomic rearrangements that alter microbial phenotype.

Integration and excision

Excision from the donor genome and integration into the recipient genome are crucial steps in the horizontal transfer of many types of mobile elements, including prophage, integron gene cassettes, and ICEs.

Site-specific integrases catalyse integration at a specific site in the host genome. For phage integrases, these are called *attB* sites. Frequently, *attB* integration sites are located within tRNA or tmRNAs, with different integrase subfamilies possessing specificity for different sublocations (such as the region coding for the anticodon loop) within the tRNA gene (Williams, 2002). Related integrase genes found on plasmids also allow site-specific integration, and share similar attachment site preferences (Williams, 2002).

A unique site-specific integrase is found in the archaea. The *Sulfolobus* spindle virus 1 (SSV1) integrase appears to recognize an *attP* site within the integrase gene itself (Table 3). This causes the integrase gene to be partitioned into an N-terminal and C-terminal fragment upon entry, apparently trapping it in the genome until it is excised by a related integrase residing on a non-integrated plasmid or virus (She *et al.*, 2004). Integrase gene fragments on either side of a region of horizontally transferred genes may therefore suggest transfer by this mechanism. However, other archaeal integrases, including the putative integrase of the plasmid pNOB8, do not appear to disrupt the gene that encodes them, so it is unclear how widespread the SSV1 integrase mechanism may be.

Integrans. Integrans are a unique class of mobile elements that utilize site-specific integration. Integrans were first identified in the late 1980s as a potentially mobile element found in association with a variety of antibiotic resistance genes (Stokes & Hall, 1989). The integron system consists of three essential elements: a tyrosine recombinase integrase gene (*intI*), a strong promoter (*pC*) and a recombination site (*attI*) (Fluit & Schmitz, 2004). Integrans do not encode the determinants for self-mobilization; rather, they facilitate the transfer of gene cassettes, which are promoterless open reading frames with *attC* (also called 59 bp) recombination sites. Specifically, the *IntI* enzyme catalyses a site-specific recombination event between the *attI* and *attC* sites, and can integrate, rearrange, or excise gene cassettes. Once incorporated into the integron, gene cassettes are expressed from the integron-based promoter *pC*; the number of integrated cassettes can range from 0 to well over 100 (Rowe-Magnus *et al.*, 2001 and references therein). Because these mobile

genes are integrated into a segment of DNA that is already capable of replication (bacterial chromosome or plasmid), and because they are integrated behind a host-compatible promoter, integrons may be important in overcoming limitations associated with the expression and replication of horizontally transferred DNA, and may therefore be an important mechanism for HGT between distantly related lineages. Finally, because integrons can be located on other mobile gene elements such as plasmids or transposons (e.g. Nesvera *et al.*, 1998), they can themselves be passed between bacteria.

There are many different classes of integrons, each class possessing a unique integrase sequence. The first integrons discovered, which contain the *intI-1* gene, were named 'type 1' integrons (Stokes & Hall, 1989). These integrons are common in pathogenic bacteria, and are one important factor contributing to the spread of antibiotic resistance (Fluit & Schmitz, 2004). Four other classes of integrons that are important in the transfer of resistance genes have been identified, and these are generally found on plasmids or transposons (Mazel, 2006). This association with mobile elements has facilitated their transfer between a diverse array of bacteria, including many species of γ -proteobacteria and Gram-positive bacteria (Fluit & Schmitz, 2004, and references therein). Integrons associated with mobile elements typically contain one to ten gene cassettes with highly variable *attC* sites.

Whole-genome sequencing has also revealed the presence of a wide variety of integrons on the chromosomes of a diverse suite of bacteria (Rowe-Magnus *et al.*, 2001, 2003; Fluit & Schmitz, 2004 and references therein; Nemergut *et al.*, 2004) (Table 3). These integrons are thought to be less likely to transfer resistance genes and to be horizontally transferred less frequently than integrons associated with other mobile elements, and they typically contain a suite of gene cassettes with homogeneous *attC* sites (Rowe-Magnus *et al.*, 2001, 2003).

Notably, although integrons have been classified according to their location (e.g. chromosomal vs mobile) or the phenotype of gene cassettes that they contain (e.g. resistance), these qualities of integrons are not stable (Stokes *et al.*, 2006), and therefore these designations should be avoided (Hall *et al.*, 2007).

Unlike other integrases, the integron integrases contain unique additional domains (see Fig. 2), which have been implicated in recombination and DNA binding (Messier & Roy, 2001 and references therein). This unique set of features is useful for detecting integron integrases in genomic or environmental data, and in distinguishing them from other tyrosine recombinases. Following detection of an integron integrase, gene cassettes can be identified using a scheme for the automated identification of gene cassette *attC* sites described by the lab of Didier Mazel (Rowe-Magnus *et al.*, 2003). Comparative analysis of the codon usages and other genomic signatures of gene cassettes, as well as sequence comparisons of gene cassette

attC sites, may yield insight into diversity or homogeneity of donors for the gene cassettes in an integron.

Homologous recombination

Horizontally transferred DNA segments that do not possess any special features that might allow integration or transposition still have a chance to be incorporated into the recipient genome. Most sequences transferred between cells by transformation, GTAs, or conjugative transfer of a chromosomal region, and many sequences transferred by generalized transduction, will depend on homologous recombination to enter the genome (Thomas & Nielsen, 2005). Rather than introducing novel DNA, homologous recombination typically replaces one DNA sequence with another, related sequence. Homologous recombination functions most efficiently when sequences are related. Thus, we should expect that these processes will be most important as mechanisms of transfer between related lineages, and they may be particularly important for promoting intraspecies orthologue displacement (Lawrence & Hendrickson, 2003).

Transposition

Although transposition per se does not allow for the intercellular transport of DNA, and thus is not, in itself, a mechanism of HGT, transposable elements facilitate gene transfer in several ways. First, intracellular transposition of genes from a microbial genome to a plasmid or phage may provide otherwise sedentary regions of DNA with a vehicle to enter other cells. Second, duplicate copies of transposable elements within a genome may aid the entry of foreign DNA carrying the same transposable elements by providing regions of homology that promote homologous recombination.

Like other mobile elements, insertion sequences (ISs) and transposons exhibit extraordinary variation both in the genes that they transfer and in their mechanism of action. ISs, the simplest form of mobile gene elements that move by transposition, consist of a transposase gene (either a serine or DDE recombinase) flanked by two inverted repeats. The transposase enzyme recognizes the inverted repeats, and catalyses recombination events that transfer the IS from one site to another within the genome. Composite transposons arise when two ISs insert near one another on a genome (e.g. Galimand *et al.*, 2005). In these cases, the outermost inverted repeats can be targets for the transposase enzyme, and the intervening sequence can be transferred along with the ISs (Fig. 2). A variety of genes have been 'trapped' on composite transposons in this fashion, including antibiotic resistance determinants (Galimand *et al.*, 2005), genes involved in metal tolerance and detoxification (Liebert *et al.*, 1999), and genes for the catabolism of xenobiotic compounds (Top & Springael, 2003). Non-composite transposons also contain a transposase and an intervening sequence, but they only contain

flanking inverted repeat sequences, and not entire ISs (e.g. Schaefer & Kahn, 1998).

ISs and transposons vary in their transposition mechanism; some move via 'replicative' transposition, while others spread via 'cut-and-paste' transposition (Curcio & Derbyshire, 2003). The cut-and-paste transposons are characterized by a transposase enzyme that excises the IS or transposon from its original site and ligates it into a new location. In contrast, replicative transposons are first copied, and the new molecule is inserted into a distal site in the genome; replicative transposons generally require the activity of both a transposase and a resolvase. However, there are mechanisms to increase the number of cut-and-paste transposons within a genome as well. For example, they can be transposed immediately after the DNA replication fork passes and be transferred to a region of the genome that has not yet been replicated.

The names and DNA sequences of many known ISs are entered into the ISFinder database, which also categorizes ISs into families (Siguier *et al.*, 2006). Further, the IslandPath program is capable of annotating transposases found near regions of putative HGT genes ('genomic islands'), which may aid in the identification of novel transposable elements (Hsiao *et al.*, 2003).

Prospects for sequence-based identification of HGT mechanisms

The information necessary to distinguish between several well-known HGT mechanisms already exists (Table 4). Published programs are available that can be used to find specific mobile elements such as prophage and transposons (Hsiao *et al.*, 2003; Bose & Barber, 2006; Fouts, 2006). An algorithm for annotating integron gene cassettes has also been developed (Rowe-Magnus *et al.*, 2003). Although no programs designed specifically to detect origins of transfer are currently available, published alignments of the various *oriT* families make the automated identification of canonical origins of transfer possible. Incorporating information about these mechanisms of transfer into algorithms for HGT detection is possible now, and could provide novel insights into how horizontal transfers inferred from genomic data were produced.

Existing programs for prophage and gene cassette annotation have proven capable of identifying mechanisms in cases where independent evidence suggests horizontal transfer. In *Salmonella typhimurium* LT2, for example, Phage_Finder was able to successfully detect the presence of all prophage in a manually curated dataset (Casjens, 2003), including the Gifsy-1, Gifsy-2, Fels-1 and Fels-2 prophage previously implicated in horizontal transfer of pathogenicity determinants amongst strains of *Salmonella enterica* (Figueroa-Bossi *et al.*, 2001). Similarly, the algorithm of Rowe-Magnus *et al.* (2003) for the detection of *attC* sites was used to compare the *attC* sites of gene cassettes that varied within the *Vibrionaceae*. These gene

cassettes were found to possess a post-segregational killing (PSK) toxin-antitoxin system more commonly associated with plasmids, which was suggested to have been acquired by transfer because of its distinct GC content compared to the genome as a whole. Thus, mechanism-based approaches have been able to detect both previously described and novel cases of gene transfer.

Differentiating between some other mechanisms of transfer is likely to be more challenging. In lineages that do not require specific DUSs for natural transformation, there may not be sequence features that can be used to annotate sites of transformation and recombination. Natural transformation followed by recombination may therefore be difficult to distinguish from other HGT processes, such as introduction of DNA by a GTA or general transducing phage followed by recombination. Similarly, until more is known, it will be difficult to develop algorithms for automatic detection of conjugation sites, like the *Streptomyces cis*-acting locus of transfer, which promote dsDNA transfer. So far, known *clt* sequences do not share a high degree of sequence similarity, although this may simply be a symptom of the few examples that are known. Further study will be necessary to determine which *clt* characteristics are best used for identification. If features of secondary structure, but not sequence, are generally conserved, then it may be necessary to make use of stochastic context-free grammars to detect and annotate those structural features (Eddy & Durbin, 1994).

Although, as outlined above, there are several promising applications for the automated annotation of mechanisms and mobile elements that are currently identifiable, other potential applications are limited by the current inability to distinguish between or detect several known mechanisms of gene transfer from sequence features alone. It would be very interesting, for example, to know how many genes in a particular genome were transferred by each mechanism. Such an analysis is currently necessarily limited to identifying genes transferred by known and identifiable mechanisms, and placing other putatively transferred genes into an 'other' category. Nonetheless, even a limited analysis of this nature may explain a substantial proportion of gene transfers. For example, Hsiao *et al.* (2005) found that 75 % of the genomic islands analysed (from a curated set of islands previously described in the literature) contained an identifiable mobility gene such as a transposase or integrase.

It should be noted, however, that the presence or absence of mobile elements or other indications of transfer within an individual genome should not be taken as representative of the strain as a whole, because horizontal transfers, or intragenomic expansions of mobile elements, can occur at the strain (e.g. Wagner, 2006) or population level (e.g. van Essen-Zandbergen *et al.*, 2007).

A further potential limitation of mechanism-based HGT studies is the presence of variant or entirely novel HGT mechanisms. We should be sceptical of the notion that all

HGT mechanisms that occur in nature have already been described – even for mechanisms of gene transfer or mobility that have been very well studied in model organisms, many variations will undoubtedly have developed over the course of microbial evolution, and these variations may make detection of a gene transfer mechanism more difficult.

Variations could include mechanisms of gene transfer that (1) make use of divergent homologues of genes involved in well-studied mechanisms, (2) involve new mobile elements comprising novel combinations of known mobility genes, (3) are variants on existing mechanisms that use homologues of known mobility genes in combination with unknown genes, and (4) are entirely novel mechanisms. Divergent homologues of known mobility genes can be detected using standard sequence similarity comparison techniques, and new mobile elements composed of novel combinations of genes encoding known transfer mechanisms can be approached using the mechanism-centric approaches to mobile elements discussed above. The techniques outlined in this review cannot directly search for entirely novel mechanisms of gene transfer. However, attempts to identify those putative transfers that do contain sequence features consistent with transfer by a well-understood mechanism could nevertheless aid efforts to uncover wholly or partially novel mechanisms of gene transfer.

Indeed, one of the great benefits of rigorous attempts to assign plausible mechanisms to horizontal transfer events is that cases will almost certainly be found in which no known mechanism can be identified. These cases could arise because: (1) no transfer occurred, but the HGT detection method inferred one anyway (false positive), or (2) HGT occurred by a known mechanism, but that mechanism could not be identified, or (3) HGT occurred by a novel, undescribed mechanism (Table 1). Careful analysis of collections of genomic regions that are predicted to have been recently transferred, but to which no known mechanism can plausibly be ascribed, could help us to better understand situations that bamboozle existing HGT detection algorithms, and may even lead to the identification of previously unknown types of mobile elements.

Mechanism annotation and global estimates of HGT frequency

One of the major remaining challenges in the area of HGT research is to achieve accurate estimates of the overall rate of HGT, and to characterize the properties of genes that are transferred most frequently. Although there are many well-established individual cases of HGT, estimation of the overall frequency of HGT on the tree of life is a difficult problem and an active area of research. Several recent attempts to estimate the global extent of HGT produced strikingly different results, probably as a result of the different methodologies used (Nakamura *et al.*, 2004; Ge *et al.*, 2005; Choi & Kim, 2007).

The large differences between the values reported using different methodologies could be due to several factors. The methods may be detecting different subsets of transfers (e.g. transfers of different ages Ragan, 2001; Ragan *et al.*, 2006), or some methods may simply be more robust to particular types of error than others. Comparison and benchmarking of the available methods against high-confidence sets of known HGT genes as well as against simulated datasets may aid understanding of the properties of these diverse methods.

Automated identification of sequence features indicative of transfer by a particular mechanism could be useful for generating the high-confidence pools of HGT genes necessary for such analyses. Such pools would also be useful because they could be used to test whether different compositional methods perform equally across identifiable transfer mechanisms. In particular, it seems possible that different compositional metrics may be particularly adept at identifying transfers facilitated by different types of mobile elements. New mechanism-specific compositional metrics for HGT detection could also be designed by training on pools of elements transferred by a similar mechanism. Finally, as discussed above, mechanism-based studies could also allow new questions to be asked about the differences, if any, between the distributions of genes transferred by various mechanisms.

Conclusion

Many difficult and exciting challenges remain to be overcome before a clearer global picture of HGT patterns can emerge. Mechanism-specific approaches, although limited by our current inability to distinguish several mechanisms of transfer using genomic data, may have the potential to leverage the enormous volume of research on diverse transfer mechanisms and mobile elements to improve our ability to detect and describe HGT events. Thus, these approaches provide a useful complement to existing compositional and phylogenetic techniques for illuminating patterns of gene transfer in the microbial world.

Acknowledgements

We would like to thank Rick Bushman, Josh Myatt, Brian Hiester, Alexey Wolfson, Micah Hamady, Catherine Lozupone, Deanna Fierman, Matt Dumlao, Erin Warddrip, Margie Krest and two anonymous reviewers for helpful comments on the manuscript. We would also like to thank the authors of many important works of primary research in the area of horizontal gene transfer that, due to space limitations, could not be cited individually. This work was supported in part by NIH predoctoral training grant T32 GM08759. J. Z. was supported by NIH predoctoral training grant T32 GM08759.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSIBLAST: a

- new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402.
- Ambur, O. H., Frye, S. A. & Tonjum, T. (2007). New functional identity for the DNA uptake sequence in transformation and its presence in transcriptional terminators. *J Bacteriol* 189, 2077–2085.
- Averhoff, B. (2004). DNA transport and natural transformation in mesophilic and thermophilic bacteria. *J Bioenerg Biomembr* 36, 25–33.
- Avery, O., MacLeod, C. & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J Exp Med* 79, 137–158.
- Bacon, D. J., Alm, R. A., Burr, D. H., Hu, L., Kopecko, D. J., Ewing, C. P., Trust, T. J. & Guerry, P. (2000). Involvement of a plasmid in virulence of *Campylobacter jejuni* 81176. *Infect Immun* 68, 4384–4390.
- Bertani, G. (1999). Transduction-like gene transfer in the methanogen *Methanococcus voltae*. *J Bacteriol* 181, 2992–3002.
- Bertani, G. & Baresi, L. (1987). Genetic transformation in the methanogen *Methanococcus voltae* PS. *J Bacteriol* 169, 2730–2738.
- Bose, M. & Barber, R. D. (2006). Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. *In Silico Biol* 6, 223–227.
- Burrus, V., Pavlovic, G., Decaris, B. & Guedon, G. (2002). Conjugative transposons: the tip of the iceberg. *Mol Microbiol* 46, 601–610.
- Bushman, F. (2002). *Lateral DNA Transfer: Mechanisms and Consequences*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L. & Brussow, H. (2003). Phage as agents of lateral gene transfer. *Curr Opin Microbiol* 6, 417–424.
- Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* 49, 277–300.
- Chen, I., Christie, P. J. & Dubnau, D. (2005). The ins and outs of DNA transfer in bacteria. *Science* 310, 1456–1460.
- Cho, E. H., Nam, C. E., Alcaraz, R., Jr & Gardner, J. F. (1999). Site-specific recombination of bacteriophage P22 does not require integration host factor. *J Bacteriol* 181, 4245–4249.
- Choi, I. G. & Kim, S. H. (2007). Global extent of horizontal gene transfer. *Proc Natl Acad Sci U S A* 104, 4489–4494.
- Court, D. L., Oppenheim, A. B. & Adhya, S. L. (2007). A new look at bacteriophage lambda genetic networks. *J Bacteriol* 189, 298–304.
- Curcio, M. J. & Derbyshire, K. M. (2003). The outs and ins of transposition: from mu to kangaroo. *Nat Rev Mol Cell Biol* 4, 865–877.
- Davidson, T., Rodland, E. A., Lagesen, K., Seeberg, E., Rognes, T. & Tonjum, T. (2004). Biased distribution of DNA uptake sequences towards genome maintenance genes. *Nucleic Acids Res* 32, 1050–1058.
- Dubnau, D. (1999). DNA uptake in bacteria. *Annu Rev Microbiol* 53, 217–244.
- Eddy, S. R. & Durbin, R. (1994). RNA sequence analysis using covariance models. *Nucleic Acids Res* 22, 2079–2088.
- Errington, J. (2001). Septation and chromosome segregation during sporulation in *Bacillus subtilis*. *Curr Opin Microbiol* 4, 660–666.
- Farahi, K., Whitman, W. B. & Kraemer, E. T. (2003). RED-T: utilizing the Ratios of Evolutionary Distances for determination of alternative phylogenetic events. *Bioinformatics* 19, 2152–2154.
- Figueroa-Bossi, N., Uzzau, S., Maloriol, D. & Bossi, L. (2001). Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Mol Microbiol* 39, 260–271.
- Fluit, A. C. & Schmitz, F. J. (2004). Resistance integrons and superintegrons. *Clin Microbiol Infect* 10, 272–288.
- Fouts, D. E. (2006). Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Res* 34, 5839–5851.
- Franco, B., Gonzalez-Ceron, G. & Servin-Gonzalez, L. (2003). Direct repeat sequences are essential for function of the *cis*-acting locus of transfer (*clt*) of *Streptomyces phaeochromogenes* plasmid pJV1. *Plasmid* 50, 242–247.
- Franke, A. E. & Clewell, D. B. (1981). Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of “conjugal” transfer in the absence of a conjugative plasmid. *J Bacteriol* 145, 494–502.
- Frost, L. S., Leplae, R., Summers, A. O. & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 3, 722–732.
- Galimand, M., Sabtcheva, S., Courvalin, P. & Lambert, T. (2005). Worldwide disseminated *armA* aminoglycoside resistance methylase gene is borne by composite transposon Tn1548. *Antimicrob Agents Chemother* 49, 2949–2953.
- Ge, F., Wang, L. S. & Kim, J. (2005). The cobweb of life revealed by genome-scale estimates of horizontal gene transfer. *PLoS Biol* 3, e316.
- Griffith, F. (1928). The significance of pneumococcal types. *J Hyg (Lond)* 27, 113–159.
- Grohmann, E., Muth, G. & Espinosa, M. (2003). Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol Rev* 67, 277–301.
- Hall, R. M., Holmes, A. J., Roy, P. H. & Stokes, H. W. (2007). What are superintegrons? *Nat Rev Microbiol* 5, C1.
- Harel, J., Duplessis, L., Kahn, J. & DuBow, M. (1990). The *cis*-acting DNA sequences required *in vivo* for bacteriophage mu helper-mediated transposition and packaging. *Arch Microbiol* 154, 67–72.
- Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D. & other authors (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406, 477–483.
- Hofreuter, D., Odenbreit, S. & Haas, R. (2001). Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Mol Microbiol* 41, 379–391.
- Hooper, S. D. & Berg, O. G. (2002). Detection of genes with atypical nucleotide sequence in microbial genomes. *J Mol Evol* 54, 365–375.
- Hsiao, W., Wan, I., Jones, S. J. & Brinkman, F. S. L. (2003). IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 19, 418–420.
- Hsiao, W. W., Ung, K., Aeschliman, D., Bryan, J., Finlay, B. B. & Brinkman, F. S. (2005). Evidence of a large novel gene pool associated with prokaryotic genomic islands. *PLoS Genet* 1, e62.
- Humphrey, S. B., Stanton, T. B., Jensen, N. S. & Zuerner, R. L. (1997). Purification and characterization of VSH1, a generalized transducing bacteriophage of *Serpulina hyodysenteriae*. *J Bacteriol* 179, 323–329.
- Jensen, E. C., Schrader, H. S., Rieland, B., Thompson, T. L., Lee, K. W., Nickerson, K. W. & Kokjohn, T. A. (1998). Prevalence of broad host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. *Appl Environ Microbiol* 64, 575–580.
- Jordan, I. K., Makarova, K. S., Spouge, J. L., Wolf, Y. I. & Koonin, E. V. (2001). Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Res* 11, 555–565.
- Karlin, S., Mrazek, J. & Campbell, A. M. (1998). Codon usages in different gene classes of the *Escherichia coli* genome. *Mol Microbiol* 29, 1341–1355.
- Kechris, K. J., Lin, J. C., Bickel, P. J. & Glazer, A. N. (2006). Quantitative exploration of the occurrence of lateral gene transfer by using nitrogen fixation genes as a case study. *Proc Natl Acad Sci U S A* 103, 9584–9589.

- Kinsella, R. J. & McInerney, J. O. (2003). Eukaryotic genes in *Mycobacterium tuberculosis*? Possible alternative explanations. *Trends Genet* **19**, 687–689.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M. & other authors (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
- Lang, A. S. & Beatty, J. T. (2007). Importance of widespread gene transfer agent genes in alpha-proteobacteria. *Trends Microbiol* **15**, 54–62.
- Lanka, E. & Wilkins, B. M. (1995). DNA processing reactions in bacterial conjugation. *Annu Rev Biochem* **64**, 141–169.
- Lawrence, J. G. & Hendrickson, H. (2003). Lateral gene transfer: when will adolescence end? *Mol Microbiol* **50**, 739–749.
- Lawrence, J. G. & Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* **44**, 383–397.
- Lawrence, J. G. & Ochman, H. (1998). Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* **95**, 9413–9417.
- Leplae, R., Hebrant, A., Wodak, S. J. & Toussaint, A. (2004). ACLAME: A CLAssification of Mobile genetic Elements. *Nucleic Acids Res* **32**, D45–D49.
- Liebert, C. A., Hall, R. M. & Summers, A. O. (1999). Transposon Tn21, flagship of the floating genome. *Microbiol Mol Biol Rev* **63**, 507–522.
- Lorenz, M. G. & Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev* **58**, 563–602.
- Marrs, B. (1974). Genetic recombination in *Rhodopseudomonas capsulata*. *Proc Natl Acad Sci U S A* **71**, 971–973.
- Mazel, D. (2006). Integrons: agents of bacterial evolution. *Nat Rev Microbiol* **4**, 608–620.
- Meibom, K. L., Blokesch, M., Dolganov, N. A., Wu, C. & Schoolnik, G. K. (2005). Chitin induces natural competence in *Vibrio cholerae*. *Science* **310**, 1824–1827.
- Messier, N. & Roy, P. H. (2001). Integron integrases possess a unique additional domain necessary for activity. *J Bacteriol* **183**, 6699–6706.
- Mirkin, B. G., Fenner, T. I., Galperin, M. Y. & Koonin, E. V. (2003). Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol Biol* **3**, 2.
- Nakamura, Y., Itoh, T., Matsuda, H. & Gojobori, T. (2004). Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* **36**, 760–766.
- Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C. & other authors (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329.
- Nemergut, D. R., Martin, A. P. & Schmidt, S. K. (2004). Integron diversity in heavy metal-contaminated mine tailings and inferences about integron evolution. *Appl Environ Microbiol* **70**, 1160–1168.
- Nesvera, J., Hochmannova, J. & Patek, M. (1998). An integron of class 1 is present on the plasmid pCG4 from gram-positive bacterium *Corynebacterium glutamicum*. *FEMS Microbiol Lett* **169**, 391–395.
- Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.
- Osborn, A. M. & Boltner, D. (2002). When phage, plasmids, and transposons collide: genomic islands, and conjugative and mobilizable transposons as a mosaic continuum. *Plasmid* **48**, 202–212.
- Page, R. D. & Charleston, M. A. (1997). From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. *Mol Phylogenet Evol* **7**, 231–240.
- Possoz, C., Ribard, C., Gagnat, J., Pernodet, J. L. & Guérineau, M. (2001). The integrative element pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer. *Mol Microbiol* **42**, 159–166.
- Prangishvili, D., Albers, S. V., Holz, I., Arnold, H. P., Stedman, K., Klein, T., Singh, H., Hiort, J., Schweier, A. & other authors (1998). Conjugation in archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. *Plasmid* **40**, 190–202.
- Ragan, M. A. (2001). On surrogate methods for detecting lateral gene transfer. *FEMS Microbiol Lett* **201**, 187–191.
- Ragan, M. A., Harlow, T. J. & Beiko, R. G. (2006). Do different surrogate methods detect lateral genetic transfer events of different relative ages? *Trends Microbiol* **14**, 4–8.
- Rowe-Magnus, D. A., Guerout, A. M., Ploncard, P., Dychinco, B., Davies, J. & Mazel, D. (2001). The evolutionary history of chromosomal superintegrations provides an ancestry for multi-resistant integrons. *Proc Natl Acad Sci U S A* **98**, 652–657.
- Rowe-Magnus, D. A., Guerout, A. M., Biskiri, L., Bouige, P. & Mazel, D. (2003). Comparative analysis of superintegrations: engineering extensive genetic diversity in the *Vibrionaceae*. *Genome Res* **13**, 428–442.
- Sato, T., Fukui, T., Atomi, H. & Imanaka, T. (2003). Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Bacteriol* **185**, 210–220.
- Schaefer, M. R. & Kahn, K. (1998). Cyanobacterial transposons Tn5469 and Tn5541 represent a novel non-composite transposon family. *J Bacteriol* **180**, 6059–6063.
- Servin-Gonzalez, L. (1996). Identification and properties of a novel *clt* locus in the *Streptomyces phaeochromogenes* plasmid pJV1. *J Bacteriol* **178**, 4323–4326.
- Sharp, P. M. & Li, W. H. (1987). The Codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**, 1281–1295.
- She, Q., Shen, B. & Chen, L. (2004). Archaeal integrases and mechanisms of gene capture. *Biochem Soc Trans* **32**, 222–226.
- Siguié, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**, D32–D36.
- Stedman, K. M., She, Q., Phan, H., Holz, I., Singh, H., Prangishvili, D., Garrett, R. & Zillig, W. (2000). pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota. *J Bacteriol* **182**, 7014–7020.
- Stokes, H. W. & Hall, R. M. (1989). A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol Microbiol* **3**, 1669–1683.
- Stokes, H. W., Nesbø, C. L., Holley, M., Bahl, M. I., Gillings, M. R. & Boucher, Y. (2006). Class 1 integrons potentially predating the association with Tn402-like transposition genes are present in a sediment microbial community. *J Bacteriol* **188**, 5722–5730.
- Syvanen, M. (1994). Horizontal gene transfer: evidence and possible consequences. *Annu Rev Genet* **28**, 237–261.
- Tatum, E. L. & Lederberg, J. (1947). Gene recombination in the bacterium *Escherichia coli*. *J Bacteriol* **53**, 673–684.
- Thomas, C. M. & Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**, 711–721.
- Tigari, S. & Moseley, B. E. B. (1980). Transformation in *Micrococcus radiodurans*, measurement of various parameters and evidence for

multiple independently segregating genomes per cell. *J Gen Microbiol* **119**, 287–296.

Top, E. M. & Springael, D. (2003). The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr Opin Biotechnol* **14**, 262–269.

Toussaint, A. & Merlin, C. (2002). Mobile elements as a combination of functional modules. *Plasmid* **47**, 26–35.

van Essen-Zandbergen, A., Smith, H., Veldman, K. & Mevius, D. (2007). Occurrence and characteristics of class 1, 2 and 3 integrons in *Escherichia coli*, *Salmonella* and *Campylobacter* spp. in the Netherlands. *J Antimicrob Chemother* **59**, 746–750.

van Passel, M. W., Bart, A., Luyf, A. C., van Kampen, A. H. & van der Ende, A. (2006). Compositional discordance between prokaryotic plasmids and host chromosomes. *BMC Genomics* **7**, 26.

Vernikos, G. S. & Parkhill, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* **22**, 2196–2203.

Wagner, A. (2006). Periodic extinctions of transposable elements in bacterial lineages: evidence from intra-genomic variation in multiple genomes. *Mol Biol Evol* **23**, 723–733.

Walsh, T. R. (2006). Combinatorial genetic evolution of multi-resistance. *Curr Opin Microbiol* **9**, 476–482.

Williams, K. P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res* **30**, 866–875.

Worrell, V. E., Nagle, D. P. J., McCarthy, D. & Eisenbraun, A. (1988). Genetic transformation system in the archaeobacterium *Methanobacterium thermoautotrophicum* Marburg. *J Bacteriol* **170**, 653–656.

Zechner, E., de la Cruz, F., Eisenbrandt, R., Grahn, A. M., Koraimann, G., Lanka, E., Muth, G., Pasegrau, W., Thomas, C. M. & other authors (2000). Conjugative DNA Transfer Processes. In *The Horizontal Gene Pool: Bacterial Plasmids and Gene Spread*, pp. 87–155. Edited by C. M. Thomas. Newark, NJ: Harwood Academic Publishers.

Zinder, N. D. & Lederberg, J. (1952). Genetic exchange in *Salmonella*. *J Bacteriol* **64**, 679–699.