

# Lateral genetic transfer: open issues

Mark A. Ragan<sup>1,2,\*</sup> and Robert G. Beiko<sup>1,2,3</sup>

<sup>1</sup>*Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia*

<sup>2</sup>*Australian Research Council Centre of Excellence in Bioinformatics, The University of Queensland, Brisbane, Queensland 4072, Australia*

<sup>3</sup>*Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia B3H 1W5, Canada*

Lateral genetic transfer (LGT) is an important adaptive force in evolution, contributing to metabolic, physiological and ecological innovation in most prokaryotes and some eukaryotes. Genomic sequences and other data have begun to illuminate the processes, mechanisms, quantitative extent and impact of LGT in diverse organisms, populations, taxa and environments; deep questions are being posed, and the provisional answers sometimes challenge existing paradigms. At the same time, there is an enhanced appreciation of the imperfections, biases and blind spots in the data and in analytical approaches. Here we identify and consider significant open questions concerning the role of LGT in genome evolution.

**Keywords:** genome evolution; lateral genetic transfer; horizontal genetic transfer; cellular networks

## 1. INTRODUCTION

Descent with modification has long been accepted as the framework within which the transmission of genetic determinants is best explained, at least in morphologically complex eukaryotes. To be sure, some aspects of eukaryotic biology lie on the edge of this paradigm (e.g. hybridization, epigenetics) or fall outside it altogether (origins of mitochondria and plastids), but these have been viewed as special cases that do not imply any fundamental problem with the standard framework. At first, it was assumed that genetic transmission in prokaryotes is fundamentally vertical too and, indeed, this explains the persistence of phenotypes over observable time, the topological congruence of gene and protein trees and many other observations. Prokaryotes, however, display a number of specialized processes unknown, or of lesser consequence, in plants or animals including transformation, transduction and conjugation, and the transfer of phenotype (e.g. antibiotic resistance or the ability to degrade xenobiotics) among strains is well known. Moreover, some gene trees are topologically discordant with each other and/or *vis-à-vis* accepted organismal relationships, and in many cases, this discordance is both statistically well supported and apparently recalcitrant to refinement of phylogenetic method. More recently, genomic sequencing has revealed patterns of gene presence and absence that cannot parsimoniously be reconciled with a purely vertical pattern of genetic transmission and gene loss. Most notably, it is now clear that in many prokaryotes, DNA can be acquired from sources other than the parent(s) and become incorporated, more or less stably, into the new genomic context. Here we offer a critical review focusing not so much on what is known about lateral genetic transfer

(LGT), but instead on what, more than a decade into the multi-genome era, remain as open issues.

## 2. PROCESS AND MECHANISM

Genomic DNA of external origin has survived release from its original host, passage through a vector and/or the environment and uptake into the new host cell; it has evaded host defences (e.g. restriction enzymes), recombined into the genome and become prevalent in a population of interest. Each of these steps constitutes a selective filter that can be of greater or lesser consequence in different recipients and under different conditions. The traditional mechanisms of conjugation, transduction and transformation (Thomas & Nielsen 2005) build on a diversity of vectors including plasmids, integrons, phage, prophage-like ‘gene transfer agents’, transposons, retrotransposons, cassette-like chromosomal elements and others that together constitute the so-called *mobilome* (Frost *et al.* 2005). These elements have probably been co-opted secondarily into LGT (Redfield 2001). Once inside the new host cell, the novel DNA may or may not replace homologous genetic material. Fluorescence microscopy can reveal LGT (Christensen *et al.* 1996, 1998) and offers broad potential for characterization of process and mechanism: fluorescent tagging indicates, for example, that most DNA transferred *via* conjugation into *Escherichia coli* is recombined into the host chromosome (Babic *et al.* 2008).

### (a) *What determines exchange partners?*

Successful conjugation or transduction requires vector compatibility between donor and recipient, which often depends on recognition of and interaction with recipient surface proteins (Thomas & Nielsen 2005). Nonetheless, conjugation between distantly related organisms such as bacteria and eukaryotes (plants and fungi) has been demonstrated experimentally

\* Author for correspondence (m.ragan@imb.uq.edu.au).

One contribution of 11 to a Theme Issue ‘The network of life: genome beginnings and evolution’.

(Heinemann & Sprague 1989); *Agrobacterium tumefaciens*, for example, can be induced to transfer its Ti plasmid to non-plant hosts including fungi and human cells in culture (Lacroix *et al.* 2006). Plasmids of the self-transmissible IncP-1 group have been isolated from both clinical and environmental (particularly wastewater) settings and harbour a wide range of resistance genes (Schlüter *et al.* 2007). These plasmids utilize mechanisms that allow their transfer, replication and maintenance in diverse Gram-negative hosts and can mobilize the transfer of other plasmids into an even wider range of target organisms. Under certain selective conditions, plasmids can expand their host range, often via a relatively small number of genetic changes (De Gelder *et al.* 2008). Broad-host-range phages are known as well: promiscuous immunoglobulin-like domains are found in three families of phages and can attach to a wide range of bacteria including *Bacillus*, *Escherichia*, *Klebsiella*, *Lactobacillus* and *Staphylococcus* (Fraser *et al.* 2006). Bioinformatic predictions of vector specificity are thus far lacking; predicting the interactions between pilus complexes and targeted surface proteins (many of which are uncharacterized) will require sophisticated structural and evolutionary models.

**(b) To what extent do different mechanisms contribute to adaptive lateral genetic transfer?**

Each step and mode of transfer exposes DNA to different types of modification. Viral DNA is subject to a much higher frequency of mutation (Drake 1991) and homologous and illegitimate recombination (Canchaya *et al.* 2003) relative to bacterial genomes. Genetic material, including physiologically important genes, carried in viral lineages for many generations can undergo radical changes (Filée *et al.* 2003; Lindell *et al.* 2004). The relative contribution of each LGT process is not well understood, but in many cases the processes leave tell-tale clues (Zaneveld *et al.* 2008). ORFans in bacteria are often enriched in A+T compared with the rest of the genome, suggesting residence in bacteriophage (Daubin & Ochman 2004). Genes on plasmids often show divergent nucleotide composition (Nakamura *et al.* 2004) and phylogenetic relationships (Gerdes *et al.* 2000) indicative of xenologous origin. Approximately 10 per cent of known bacterial genomes harbour one or more integrons (Boucher *et al.* 2007), which appear to be more mobilizable and of broader phyletic distribution than recognized until recently. The presence of transposon or integrase sequences suggests past mobility (Vernikos & Parkhill 2008). A particularly long integrated region (e.g. tens of thousands of base pairs) may constitute evidence for transduction rather than conjugation or transformation, whereas short regions do not distinguish among potential mechanisms because not all introgressed DNA may have been recombined into the host genome, or preserved thereafter. New insights and approaches are required to deal with mechanisms that leave no trace (other than the integrated genetic material), with unknown mechanisms, and with mutational decay of characteristic sequences.

**(c) How does environment constrain lateral genetic transfer?**

Natural environments (e.g. soil) can potentially be conducive to LGT. Concentrations of free DNA can exceed 1 µg/g (Niemeyer & Gessler 2002; Vlassov *et al.* 2007), and under some conditions, DNA can persist for millennia (Austin *et al.* 1997) although its bioactivity may be modulated, e.g. by nucleases or humic acids. Some 90 species of prokaryotes, including some that live in soil or water, are known to be naturally competent (Nielsen *et al.* 1998; de Vries & Wackernagel 2005). Potential hot spots in the environmental for recombination and LGT include biofilms, the rhizosphere, decomposing material, guts of soil animals and within bacterivorous protozoa. The nature, frequency and determinants of LGT in these complex environments are beginning to be approached via metagenomic sequencing.

It has been assumed that a key limitation on the transfer of genetic material is the need for donor and recipient to be present either in the same habitat, or in two habitats that are linked by a biotic or an abiotic 'bridge'. Generalist organisms such as *Pseudomonas aeruginosa* have relatively large genomes and can survive in a wide range of habitats, readily expanding and contracting their genetic content to facilitate new adaptations (Schmidt *et al.* 1996). A striking example of genomic plasticity is the recent identification in a clinical strain of *P. aeruginosa* of genes for degradation of secondary metabolites of trees (Mathee *et al.* 2008). Environmental ranges may have been underestimated: *Chlamydiae* have recently been discovered inside amoebae and in wastewater (Horn & Wagner 2001) and *Rickettsia* are viable in amoebae (Ogata *et al.* 2006), both likely LGT hot spots. Viruses are capable of moving between habitats (Breitbart & Rohwer 2005). Can DNA be exchanged, perhaps inefficiently, across the entire biosphere?

Intracellular bacteria have been thought to be largely unaffected by LGT, due to their isolation from potential donors (Renesto *et al.* 2005). Prophages, plasmids and transposons (Bordenstein & Reznikoff 2005) are, however, found in many intracellular bacteria including *Buchnera*, which displays no evidence of LGT. Conjugative plasmids occur in the intracellular parasite *Rickettsia felis* (Ogata *et al.* 2005) and may be important in the evolution of virulence (Gillespie *et al.* 2007). There is evidence for extensive between-strain recombination in the intracellular parasite *Chlamydia trachomatis*, although many recombined sequences are too long to have been acquired in single transformation events, and no potential vector has been identified (DeMars & Weinfurter 2008). Given that exchange is possible, are the constraints to LGT mainly environmental (e.g. lack of suitable genes, selective pressure for small genome size) or internal (configuration of the reduced cellular network) (Jain *et al.* 2003)?

**(d) What is the physical unit of transfer?**

Lateral origins have been proposed for DNA regions ranging from seven nucleotides (Denamur *et al.* 2000) up to an entire chromosome greater than 3 Mb

(Lin *et al.* 2008) in length and constituting stretches of non-coding DNA, portions of genes, intact genes, multi-gene clusters, operons, plasmids, transposable elements and pathogenicity islands. For this reason, we refer to lateral (or horizontal) *genetic*, not *gene*, transfer. As it is doubtful that a Mb-sized region could be successfully transferred except by cell fusion, or integrated all at once into the host regulatory network, these largest regions are presumably built up via a succession of transfers, either primary LGT events targeting a receptive area or selection-driven reshuffling of originally dispersed elements (van der Does & Rep 2007).

Would whole-gene LGT be selectively most advantageous? In the immediate term, especially where the introgressed gene does not require efficient regulation and the protein is functional without having to be part of a complex, whole- and multi-gene LGT offer immediate functionality (e.g. antibiotic resistance). On the other hand, significant up-front costs to competitive fitness might be incurred before amelioration and recruitment of regulators eventually allow efficient stoichiometric participation in cellular networks; these costs, however, might be lessened by the silencing or transcriptional down-regulation of incoming genes, whether actively by the H-NS system in Gram-negative bacteria (Dorman 2007) or passively by regulator and codon-profile mismatch.

Homologous replacement of part of a gene (within-gene LGT) might avoid some of these problems but raises others, including possible disruption of the folding, catalysis or interactions of the chimaeric gene product. Nonetheless, some examples have been identified in prokaryotes, including within the *rrnB* operon (Yap *et al.* 1999) and the *mutU* and *mutS* genes (Denamur *et al.* 2000); many others are implicit from genome-wide scans of recombination breakpoints (Mau *et al.* 2006). Of two instances of LGT within genes encoding EF-1 $\alpha$  (Inagaki *et al.* 2006), one preserves existing domains but the other interrupts domains. Are domains (sometimes) the unit of LGT? Chan *et al.* (2009) found that 25.0 per cent of 1462 sets of orthologous genes exhibit at least one within-gene recombination breakpoint. These breakpoints are not uniformly distributed among gene sets, but preferentially occur in sets with annotated SCOP domains and often interrupt these domains. Thus, DNA regions that encode full-length protein domains (which we might term *domons*) are not privileged units of LGT. Many genes have mixed heritage, and greater precision is needed in describing them as *vertical* or *lateral*, *concordant* or *discordant*.

### 3. QUANTIFICATION

There is a broad consensus that LGT has contributed to genome evolution in prokaryotes (and to some extent in eukaryotes: see below). There is much less agreement on its quantitative extent: estimates range from close to zero, to more than one LGT event per gene per genome. At what point do lateral signatures obscure the vertical, moving LGT from an important but secondary player to the centre stage of genome evolution?

#### (a) *How should we think about and express the extent of lateral genetic transfer?*

In the phylogenetic approach, each instance of topological discordance between a gene tree and a trusted reference tree is taken as a *prima facie* instance of LGT. Discordance can be found throughout the entire range of nodal depths within these trees, from recent (genera, species) to older, presumably reflecting a commerce in genetic material that has been ongoing since pre-genomic times (Woese 2000). Viewed in this way, every genome has LGT in its ancestry. This very situation, however, implies that whatever meaning there may be in the concept *genome phylogeny* does not extend simply to each constituent gene or region. If we instead ask what proportion of genes (or other analysable units) in a genome exhibit, back to the deepest defined node, a path that is discordant with a reference hypothesis (e.g. a reference topology, or monophyly of a particular taxon), the number can be 25–50% or more (Nelson *et al.* 1999; Kunin & Ouzounis 2003; Creevey *et al.* 2004; Lerat *et al.* 2005; Zhaxybayeva *et al.* 2006; Dagan & Martin 2007; Shi & Falkowski 2008). But this approach gives multiple votes to basal edges, which lie on many paths and thus may be traversed multiple times, without necessarily taking into account the concordance of subtrees and, in most implementations, the support for individual edges. In an extreme case, all genes descending from a single ancient transfer event might be counted as lateral, even if every subtended node and edge agrees perfectly with the reference tree.

Thus, a third approach is to focus on the proportions of recognizable genetic transmission steps (edges) that are vertical or horizontal, i.e. on the proportion of well-supported bipartitions that are discordant with a reference hypothesis. Beiko *et al.* (2005) identified 13.4 per cent of nearly 100 000 bipartitions, over 144 prokaryotic genomes, as concordant at posterior probability (PP)  $\geq 0.95$  with a reference supertree; analysing a 165-genome dataset, Kunin *et al.* (2005) found 4.7–5.2% of events to be LGT, 11.1–11.6% gene losses and 83.4–83.6% vertical transfers. At an even finer level of resolution, each edge represents many millions of organismal generations but, even so, many internal edges remain consistent with reference tree: thus instantaneously, vertical events (which, however, cannot be distinguished from LGT among very close relatives) are vastly more likely to persist than are lateral transfers.

Each of these approaches addresses a distinct question and yields a distinct perspective on the contribution of LGT to genome evolution. More generally, we might ask whether it is better to *count* (inferred) transfers or to *model* transfer. A quantitative model would attempt to account for events that are not empirically detectable by comparative approaches (e.g. due to their transitory existence, near-identity of sequence or obliteration by subsequent events) or fall short of a defined statistical threshold. Promising steps have been taken in Bayesian (Suchard 2005) and likelihood (Galtier 2007) frameworks; as further genomes are sequenced, it will be possible to estimate parameter values for models of increasing complexity.

**(b) Can phylogenetic and non-phylogenetic approaches provide complementary evidence?**

Diverse problems can afflict tree-based approaches: non-robust or inconsistent methods, model violation, lack of convergence in Markov chain Monte Carlo sampling, paralogy, ancient lineage-sorting, inadequate taxon sampling and others. Over 22 432 gene trees, Beiko *et al.* (2005) showed that the underlying data offer the same statistical support to instances of topological incongruence (i.e. LGT) as to congruence (vertical descent), but deeper systematic problems may remain hidden. Non-phylogenetic (*surrogate*) methods may complement the phylogenetic approach: a strongly G+C-biased gene, for example, might violate phylogenetic models but be readily detected by its anomalous composition. Surrogate methods differentially detect LGT of different relative ages (Ragan *et al.* 2006) and could be chosen specifically to complement weakly supported nodes. The applicability of surrogate data may be case specific: gene (DNA tract) order is much better conserved in *Yersinia* (Darling *et al.* 2008), for example, than in *Escherichia*. But in general, does the union of all prediction sets represent the best estimate of total LGT (Lawrence & Ochman 2002; Ragan 2002)?

Flanking phage integrase genes or integration motifs are often taken as 'smoking gun' evidence of LGT in prokaryote genomes. Cho & Palmer (1999) made an analogous argument for co-conversion tracts flanking a group I intron in angiosperm mitochondria. Repeating the analysis with more-extensive taxon sampling, Cusimano *et al.* (2008) concluded that intron distribution is best explained by vertical transmission plus multiple losses and, hence in this case at least, co-conversion tract footprints are unreliable markers for LGT.

**(c) How much lateral gene transfer is undetectable?**

Methodological approaches are differentially blind to LGT. Methods based on atypical composition or hidden Markov models may not detect ancient, well-ameliorated transfers (Ragan *et al.* 2006); gene distributions brought about by LGT may be mistakenly attributed to gene loss (and vice versa); transfers between sister leaves are phylogenetically silent. Poptsova & Gogarten (2007) compared three tests used in conjunction with phylogenetic approaches: the 'approximately unbiased' (AU) test, the Robinson–Foulds symmetric distance and bipartition spectra. For *in silico* 'transfers' superimposed on the terminal branches of a  $\gamma$ -proteobacterial phylogeny, bipartition spectra failed to detect only 3 per cent and 6 per cent of transfers at 70 per cent and 90 per cent cut-offs, while generating fewer than 4 per cent and 2 per cent false positives, respectively. The AU test was less sensitive (and requires knowledge of the true tree), whereas at 97.5 per cent cut-off the Robinson–Foulds distance detected only 58 per cent of reciprocal exchanges and 60 per cent of orthologous replacements. The inability of most approaches to detect transfers among sister leaves is particularly problematic, as diverging strains are exactly the ones most likely to

live in similar environments, possess similar gene contents and genomic tracts amenable to homologous recombination, and harbour the same vectors; on the other hand, most LGT between sisters may be transitory 'churn' (futile turnover) that is best ignored in most contexts.

**(d) Can aggregate methods provide a reliable reference tree?**

By integrating over diverse genome regions and cellular functions, aggregate approaches may avoid biases inherent in simple hypotheses or individual approaches, particularly if data giving rise to discordant signals are first removed. Aggregation might be accomplished via a supermatrix of concatenated sequence data or a supertree computed from well-supported features of individual gene trees, or be captured in an overall distance measure based on gene content, order or pairwise similarity. But is the resulting aggregate signal that of a vertical phylogenomic core? Two objections have been raised: that this is philosophically the wrong view of genome evolution and that vertical signal may be obscured by LGT. The former is considered in two other papers in this issue (Doolittle 2009; Fournier *et al.* 2009); the latter objection can be approached via computational simulation. Ge *et al.* (2005) found that even at several LGT events per gene family, the vertical backbone can be identified clearly. Based on more-extensive simulations, Beiko *et al.* (2008) show that vertical signal is preserved even at high LGT frequencies except in the presence of biases (e.g. habitat-directed bias) that favour exchange among more distantly related genomes. When such biases are present, aggregate histories reflect neither the true vertical history nor the lateral alternative. Supertree methods may be less susceptible to some LGT biases (Gribaldo & Brochier-Armanet 2006).

**(e) Can the biological sources of transfers be identified?**

LGT can be modelled as a break-and-reanneal operation on a graph. A gene (together with the subtree of its descendants) is broken off the reference (genomic or organismal) tree and reannealed at a topologically different position on its gene tree; the initial breakage point thus represents its ultimate biological source. This same point, however, also represents all the extinct, undiscovered and un-sequenced lineages that would, if known, descend from that edge (Gogarten & Townsend 2005), so the reference tree itself may tell us little about the biology of the primary donor. Where multiple break-and-reanneal steps are required, a shortest path through intermediate host lineages can often be inferred (Beiko & Hamilton 2006), but there is no guarantee that LGT has taken the shortest computational path: the intermediates may likewise be extinct, unknown or un-sequenced, or the path may be ecologically improbable. Kunin *et al.* (2005) propose that 'hubs' in LGT networks serve as 'gene banks' that acquire and redistribute genes within microbial communities; as above, these hubs too may represent extinct, unknown and un-sequenced organisms.

Composition-based methods offer even lower resolution. As described above, DNA of lateral origin has passed filters that can skew its composition. In most  $\gamma$ -proteobacteria, ORFans (genes without significant matches in current databases) are A+T-rich, suggesting a phage ancestry (Daubin *et al.* 2003; Daubin & Ochman 2004). Successive LGT events can superimpose DNA onto previously incorporated regions, resulting in a genomic pastiche that may defy analysis; Chan *et al.* (2009) found that 5.5 per cent of 1462 orthologue families exhibit complex recombination patterns consistent with successive layering of LGT. Gene content can be useful: a gene showing sparse, anomalous presence in one lineage might be inferred to have been transferred from a group in which it is ubiquitous.

If LGT has been frequent, then physiology (hence niche) has presumably changed repeatedly over time. A probable example is the plant pathogen *Erwinia carotovora* subsp. *atroseptica*, which shares the common enterobacterial backbone but has gained many specializations for its plant-pathogenic lifestyle via LGT from other bacteria living 'in and around plants' (Toth *et al.* 2006).

#### (f) *Is lateral gene transfer less frequent among eukaryotes?*

DNA appears to be readily transferred from intracellular prokaryotes to the eukaryotic host nucleus: 18 per cent of *Arabidopsis* nuclear genes may be of cyanobacterial, presumably proto-plastid, origin (Martin *et al.* 2002) and much of *Wolbachia* is found in its arthropod and nematode host nuclei. In *Drosophila*, 2 per cent of the transferred *Wolbachia* genes are expressed (Dunning Hotopp *et al.* 2007). Plant-associated nematodes appear to have acquired cellulases and pectinases from plant-associated bacteria (Mitrevva *et al.* 2005), while genes of probable prokaryotic origin are responsible for cellulose synthesis (Nakashima *et al.* 2004), starch degradation (Da Lage *et al.* 2007), shikimate biosynthesis (Richards *et al.* 2006) and detoxification (Burroughs *et al.* 2006) in various animals. Gene flow in the opposite direction appears to be less common, although possible examples have been put forward (Ponting *et al.* 1999; Jenkins *et al.* 2002; Chen *et al.* 2007).

LGT is well documented among fungi (Oliver & Solomon 2008). For example, an 11-kb DNA region that includes a gene (*ToxA*) for toxin production and a transposase has moved from the nuclear genome of one wheat pathogen, *Pyrenophora tritici-repentis*, to that of another, *Stagonospora nodorum* (Friesen *et al.* 2006). Plant mitochondria, although not plastids, exchange genes across species (Richardson & Palmer 2007) and might serve as portals into plant nuclear genomes. On the other hand, LGT is infrequent in yeasts and animals (Andersson 2005), although Burghoff *et al.* (2008) provide evidence for transfer of the enhanced green fluorescent protein (EGFP) marker from human endothelial cells to rat cardiomyocytes *via* apoptotic bodies after intra-coronary transplantation; they found no evidence for nanotubular connections between cells, nor for a viral vector. It is not known whether the low frequency of LGT among animals is primarily physical (protection of

the germ line from intracellular bacteria) or regulatory (difficulty in integrating into a complex regulatory network, e.g. involving non-coding RNAs).

Most eukaryotic diversity lies among protists. Relatively few protist nuclear genomes have been sequenced, but limited data suggest that phagotrophic protists may have LGT frequencies comparable to those of bacteria (Richards *et al.* 2003; Andersson 2005).

#### 4. IMPACT

It is well known that genes encoding uptake, transport and metabolic functions (which can convey selective advantage in certain environments) have often been transferred laterally: examples include bacterial transporters (Gelfand & Rodionov 2008), nitrogen fixation (Raymond *et al.* 2004), type III secretory systems (Tobe *et al.* 2006) and many others. But no category of gene is immune to LGT: even informational genes involved in DNA mismatch repair (Lin *et al.* 2007) and translation elongation (Inagaki *et al.* 2006) show strong evidence for transfer. Like gene duplication, LGT can provide fodder for neofunctionalisation: for instance, the type IV secretion system in *Bartonella* appears to have been acquired via LGT, following which its function changed from conjugation to erythrocyte adherence (Nystedt *et al.* 2008). With the help of mobile genetic elements, novel pathways can be assembled by LGT (Springael & Top 2004).

#### (a) *How long do laterally transferred sequences persist in a genome?*

Individual genome sequences are snapshots of evolution: some genes are well integrated and will presumably remain in the lineage for a long term, while others are transient and unlikely to persist. The highest rates of gene gain are found at the leaves of the tree, implying that many genes gained earlier were subsequently lost (Berg & Kurland 2002; Hao & Golding 2006). Not all lateral genes, however, turn over rapidly; in the  $\gamma$ -proteobacteria, for example, some persist vertically for great lengths of time and even become characteristic of the group (Lerat *et al.* 2003). At least in *E. coli/Shigella*, genes acquired from distantly related bacterial groups are less likely to persist than ORFans (von Passel *et al.* 2008). Eukaryotic groups too may be defined by lateral genes that were fixed in a common ancestor: red algae and green plants, for instance, uniquely among eukaryotes share *TOP6B*, a gene of archaeal origin (Huang & Gogarten 2006). Lima *et al.* (2008) classified genomic islands of *Xanthomonas* into two categories based on atypicality of nucleotide compositions; deviant islands, which potentially have been acquired more recently, are rich in mobile genetic elements, whereas islands of typical (ameliorated?) composition harbour many more genes involved in metabolism and cellular processes.

#### (b) *How are lateral genes connected to cellular networks?*

The workings of the cell can be abstracted as a network in which molecules (genes or proteins) are represented as nodes (vertices) and physical interactions as edges. Given the state of knowledge, these representations

contain false-positive and false-negative edges and ignore spatial and temporal details. Nonetheless, important generalizations emerge: (i) cellular networks contain both highly and weakly connected nodes; (ii) a species-specific *core* subset of nodes is present in all strains of a species, while other nodes may be found in most, some or only a few strains; (iii) core nodes are chromosomal, whereas peripheral nodes may be encoded on the chromosome, islands or plasmids; (iv) core nodes typically describe functional units, e.g. operons or macromolecular complexes; (v) core nodes tend to be more highly connected and more highly expressed than peripheral nodes; (vi) core-node genes accumulate point mutations more slowly than do peripheral-node genes; (vii) peripheral nodes are more often implicated in functions that are directly affected by the environment and, at least in human, the periphery of the protein-interaction network maps approximately to the cellular periphery; and (viii) networks evolve by the addition of peripheral nodes (Pál *et al.* 2005; Kim *et al.* 2007; Lercher & Pál 2008; Wellner *et al.* 2007; Davids & Zhang 2008). These features of biological networks contribute to their evolvability (Oikonomou & Cluzel 2006).

Because they are highly connected, require tighter stoichiometric and expression control, are less-exposed to immediate selective pressure or are evolutionarily conservative, core-node genes are less susceptible to homologous replacement via LGT than their peripheral counterparts (Jain *et al.* 1999; Papp *et al.* 2003; Aris-Brosou 2005). At least, this appears true for lateral genes identified by compositional bias, i.e. those more recently arrived in the genome, although not for those identified by tree comparison (Wellner *et al.* 2007). Under a generous definition, the *Streptococcus* core genome is about 18 per cent recombinant while the *Streptococcus pyogenes* core may be 35 per cent recombinant. Thus, lateral genes that survive the initial 'churn' can slowly become better integrated into host-cell networks, e.g. by recruiting transcriptional regulators (Navarre *et al.* 2007; Wellner *et al.* 2007; Lercher & Pál 2008), recognizing transcription, translation, folding and assembly signals (Lercher & Pál 2008), and otherwise accommodating kinetically and thermodynamically. Most global regulators have evolved vertically, whereas many local regulatory nodes have been acquired by LGT, often simultaneously with the gene(s) they regulate (Price *et al.* 2008). Interestingly, in both *E. coli* and yeast, lateral genes preferentially interact with core nodes (Eisenberg & Levanon 2003; Davids & Zhang 2008).

Network remodelling can occur on large scale, e.g. in invasion of new niche: lactobacilli making the transition to life in a nutritionally rich medium lose 600–1200 genes and gain almost 100 others perhaps by LGT (Makarova *et al.* 2006). Except in the case of extreme genomic reduction, cellular networks remain evolvable, yet robust to the vicissitudes of short-term LGT.

#### (c) *Do all prokaryotes have a stable core genome?*

The concept of a core genome (as in the preceding section) is interrelated with the delineation of groups of

organisms. Groups can be delineated, perhaps hierarchically, based on sequence hybridization or identity (Stackebrandt & Goebel 1994), population-level processes (Cohan 2002) or a hybrid of genomics and ecology (Staley 2006). A set of universally (or nearly universally) conserved genes may define each group. Among eight *Streptococcus agalactiae* genomes, the core was estimated to be approximately 90 per cent as large as the smallest individual genome (Tettelin *et al.* 2005), whereas in *E. coli*, the core genome of the first three sequenced strains was about 70 per cent the size of *E. coli* K12, demonstrating much higher variation in gene content (Welch *et al.* 2002). In subsequent analysis of 20 sequenced *E. coli* strains, the use of a slightly relaxed core criterion drastically increases the estimated size of the core (Charlebois & Doolittle 2004; Konstantinidis *et al.* 2006). The relationship between strict and ancestral cores should be explored further, to understand why genes sufficiently important to remain in most organisms can be lost in a small subset.

One (perhaps the dominant) genomic component of a species is its 'clonal frame', ancestrally inherited genetic material with conserved linkage. This clonal frame may, however, be progressively disrupted as genes are introduced via LGT from outside the clonal population and through other events such as gene loss (Milkman & Bridges 1990). Different species and populations show different levels of clonality: *Salmonella enterica*, for example, shows high levels of clonality and linkage disequilibrium, whereas *Neisseria gonorrhoeae* exhibits a much higher rate of mixing and almost no clonal frame (Sriramulu 2008). Linkage equilibrium has been observed in natural populations such as those of *Halorubrum* (Papke *et al.* 2004). The core can also include genes that were introduced more recently than the common species ancestor and have spread through the population via LGT; these components should exhibit different evolutionary histories relative to the clonal frame and may not show conserved linkage relative to the clonal frame in different strains or isolates. Most analyses of clonal frame have sampled a few housekeeping loci from many isolates of a given species rather than surveying entire genomes. Such analyses have revealed, for instance, contrasts in the population structures of environmental versus pathogenic members of a species (e.g. Bisharat *et al.* 2007). Empirical studies of the stability of clonal frames within or outside the taxonomic group of interest, and of the rate and mode of clonal frame decay at different phylogenetic depths, are, however, thus far lacking. The existence of a clonal frame might provide a strong hypothesis for organismal relationships among members of a species, while acquired core genes may point to functions that are sufficiently valuable to have become fixed.

#### (d) *What factors determine the size and composition of pan-genomes?*

Sonea & Panisset (1983) envisioned all prokaryotes as forming a global gene pool, from within which different subsets of genes come together in unlimited, if temporary, combinations. This concept has been

refined to the idea of the *pan-genome*, the entire repertoire of genetic information in all genomes of a given species (Tettelin *et al.* 2005). Estimates of pan-genome size are affected by non-random strain sampling and extrapolation effects, but in some cases reach an order of magnitude greater than the size of any individual genome (Medini *et al.* 2005). In other cases (e.g. *Bacillus anthracis*), the pan-genome appears closed, and a few genomes yield a nearly complete inventory. Habitat range appears to be important in determining pan-genome size, with the broad-range *S. agalactiae* having a much larger pan-genome than *S. pyogenes*, primarily found in nasal mucosa (Lefebvre & Stanhope 2007). Susceptibility to different mechanisms of LGT may also play a role in determining pan-genome size.

Genes that have recently joined the *accessory genome* may reflect current and recent environmental stresses experienced by different isolates within a species (Thomason & Read 2006). This hypothesis is borne out by many analyses of core versus accessory genomes. A comparative analysis of 12 strains of *Prochlorococcus marinus* revealed conserved core genes, sets of genes that are ubiquitous in and exclusive to high-light- and low-light-adapted subsets of the group, plus accessory genes specific to one or a few strains (Kettler *et al.* 2007); strain-restricted functions include viral defence and transport of toxins. Association of adaptive genes by ecotype or lifestyle is compelling, but candidate virulence genes in *Neisseria* (Schoen *et al.* 2008) and *Streptococcus* (McMillan *et al.* 2006) do not preferentially associate with virulent strains, suggesting that sequence variation can be important. Eppinger *et al.* (2004) identified a 'flexible gene pool' for four *Campylobacter* species comprising plasmids, phage elements, mobile DNA and genomic islets and DNA-restriction/modification systems. Elucidating pan-genome function and possible modularity of different components of the pan-genome will be difficult, particularly since many non-core genes are conserved hypotheticals and ORFans. Functional predictions will need to be complemented with mutagenesis and gene deletion experiments, to assess the contribution of different pan-genomic elements to organism and community function.

#### (e) *Can we reconstruct ancestral physiologies?*

Dagan & Martin (2007) used a parsimony approach to infer ancestral genome sizes under different frequencies of LGT, observing that an average of 1.1 LGT events per genome family yielded ancestral genomes of a similar size distribution to modern ones. However, if larger genomes with generalist lifestyles are more likely to produce descendants that persist in the long term, then ancestral genomes may have been larger, on average, than modern ones and LGT frequencies of 0.5–1.0 per gene family may not be unrealistic. This in turn might be counterbalanced to some extent if ancient enzymes were, on average, more broadly versatile than their modern descendants (Cairns-Smith & Walker 1974; Yčas 1974).

Beyond the question of ancestral genome sizes, ancestral reconstructions of gene content and

metabolic potential are useful in understanding the evolutionary pathways followed by organisms and inferring the features of ancient habitats. Parsimony analyses implicitly generate ancestral genomes at the internal nodes of a tree or network (Ouzounis 2005). Using reference genome trees and phylogenetic profiles, Shi & Falkowski (2008) extrapolated that the last common ancestor of cyanobacteria was incapable of nitrogen fixation and was likely thermophilic. Core informational and photosynthetic functions were probably present in this ancestor, while some peripheral components of the photosynthetic apparatus may have been absent. Mapping core genes into an inferred common ancestor is parsimonious, but like present-day prokaryotes this ancestor presumably contained variable-shell genes that have not persisted. In general, the accuracy with which we can reconstruct an ancestral genome will be affected by the distribution of duplication, loss and LGT events as well as by the completeness of descendant taxon sampling. Biased gene loss or gene uptake (e.g. niche specialization or colonization of new habitats) is likely to confound reconstruction: it is unlikely that the genome of the last free-living ancestor of endosymbionts such as *Buchnera* and *Wigglesworthia* can be inferred from present-day symbiont genomes. Reconstructed ancestral genomes must also be coherent biochemically and energetically: a patchwork of fragmentary pathways is unlikely to be accurate. Finally, the underlying notion of an ancestral genome loses meaning if LGT has been sufficiently frequent or the inferred ancestor sufficiently ancient: in such cases, the modern genome will represent an amalgam of multiple ancestral lineages, perhaps without preference for the genome of the cellular ancestor.

## 5. PROSPECTUS

Great progress has been made in understanding LGT as a lifestyle for populations and species, but key issues—including those identified above—remain open. We offer this brief overview in hopes of stimulating the formulation of hypotheses that can be informed by new data, are susceptible to experiment or can be sharpened by mathematical modelling. A deeper understanding of the dynamics of genome evolution, including LGT, will inform scientific and public debate on a number of critical issues including biodiversity, global climate change and the engineering of biological systems.

We acknowledge the support of Australian Research Council (grant CE0348221) to M.A.R. and from the Canada Research Chairs program, Genome Atlantic and the Natural Sciences and Engineering Research Council to R.G.B.

## REFERENCES

- Andersson, J. O. 2005 Lateral gene transfer in eukaryotes. *Cell. Mol. Life Sci.* **62**, 1182–1197. (doi:10.1007/s00018-005-4539-z)
- Aris-Brosou, S. 2005 Determinants of adaptive evolution at the molecular level: the extended complexity hypothesis. *Mol. Biol. Evol.* **22**, 200–209. (doi:10.1093/molbev/msi006)

- Austin, J. J., Smith, A. B. & Thomas, R. H. 1997 Palaeontology in a molecular world: the search for authentic ancient DNA. *Trends Ecol. Evol.* **12**, 303–306. (doi:10.1016/S0169-5347(97)01102-6)
- Babic, A., Lindner, A. B., Vulic, M., Stewart, E. J. & Radman, M. 2008 Direct visualization of horizontal gene transfer. *Science* **319**, 1533–1536. (doi:10.1126/science.1153498)
- Beiko, R. G. & Hamilton, N. 2006 Phylogenetic identification of lateral genetic transfer events. *BMC Evol. Biol.* **6**, 15. (doi:10.1186/1471-2148-6-15)
- Beiko, R. G., Harlow, T. J. & Ragan, M. A. 2005 Highways of gene sharing in prokaryotes. *Proc. Natl Acad. Sci. USA* **102**, 14 332–14 337. (doi:10.1073/pnas.0504068102)
- Beiko, R. G., Doolittle, W. F. & Charlebois, R. L. 2008 The impact of reticulate evolution on genome phylogeny. *Syst. Biol.* **57**, 844–856. (doi:10.1080/10635150802559265)
- Berg, O. G. & Kurland, C. G. 2002 Evolution of microbial genomes: sequence acquisition and loss. *Mol. Biol. Evol.* **19**, 2265–2276.
- Bisharat, N., Cohen, D. I., Maiden, M. C., Crook, D. W., Peto, T. & Harding, R. M. 2007 The evolution of genetic structure in the marine pathogen, *Vibrio vulnificus*. *Infect. Genet. Evol.* **7**, 685–693. (doi:10.1016/j.meegid.2007.07.007)
- Bordenstein, S. R. & Reznikoff, W. S. 2005 Mobile DNA in obligate intracellular bacteria. *Nat. Rev. Microbiol.* **3**, 688–699. (doi:10.1038/nrmicro1233)
- Boucher, Y., Labbate, M., Koenig, J. E. & Stokes, H. W. 2007 Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol.* **15**, 301–309. (doi:10.1016/j.tim.2007.05.004)
- Breitbart, M. & Rohwer, F. 2005 Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **13**, 278–284. (doi:10.1016/j.tim.2005.04.003)
- Burghoff, S. *et al.* 2008 Horizontal gene transfer from human endothelial cells to rat cardiomyocytes after intracoronary transplantation. *Cardiovasc. Res.* **77**, 534–543. (doi:10.1093/cvr/cvm071)
- Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D. & Aravind, L. 2006 Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphotesterases and allied enzymes. *J. Mol. Biol.* **361**, 1003–1034. (doi:10.1016/j.jmb.2006.06.049)
- Cairns-Smith, A. G. & Walker, G. L. 1974 Primitive metabolism. *BioSystems* **5**, 173–186. (doi:10.1016/0303-2647(74)90030-6)
- Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.-L. & Brüssow, H. 2003 Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* **6**, 417–424. (doi:10.1016/S1369-5274(03)00086-9)
- Chan, C., Darling, A. E., Beiko, R. G. & Ragan, M. A. 2009 Are protein domains modules of lateral genetic transfer? *PLoS ONE* **4**, e4524. (doi:10.1371/journal.pone.0004524)
- Charlebois, R. L. & Doolittle, W. F. 2004 Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res.* **14**, 2469–2477. (doi:10.1101/gr.3024704)
- Chen, L.-L., Wang, G.-Z. & Zhang, Y. 2007 Sterol biosynthesis and prokaryotes-to-eukaryotes evolution. *Biochem. Biophys. Res. Commun.* **363**, 885–888. (doi:10.1016/j.bbrc.2007.09.093)
- Cho, Y. & Palmer, J. D. 1999 Multiple acquisitions *via* horizontal transfer of a group I intron in the mitochondrial *cox1* gene during evolution of the Araceae family. *Mol. Biol. Evol.* **16**, 1155–1165.
- Christensen, B. B., Sternberg, C. & Molin, S. 1996 Bacterial plasmid conjugation on semi-solid surfaces monitored with the green fluorescent protein (GFP) from *Aequorea victoria* as a marker. *Gene* **173**, 59–65. (doi:10.1016/0378-1119(95)00707-5)
- Christensen, B. B., Sternberg, C., Andersen, J. B., Eberl, L., Moller, S., Givskov, M. & Molin, S. 1998 Establishment of new genetic traits in a microbial biofilm community. *Appl. Environ. Microbiol.* **64**, 2247–2255.
- Cohan, F. M. 2002 What are bacterial species? *Annu. Rev. Microbiol.* **56**, 457–487. (doi:10.1146/annurev.micro.56.012302.160634)
- Creevey, C. J., Fitzpatrick, D. A., Philip, G. K., Kinsella, R. J., O'Connell, M. J., Pentony, M. M., Travers, S. A., Wilkinson, M. & McInerney, J. O. 2004 Does a tree-like phylogeny only exist at the tips in the prokaryotes? *Proc. R. Soc. B* **271**, 2551–2558. (doi:10.1098/rspb.2004.2864)
- Cusimano, N., Zhang, L.-B. & Renner, S. S. 2008 Reevaluation of the *cox1* group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. *Mol. Biol. Evol.* **25**, 265–276. (doi:10.1093/molbev/msm241)
- Da Lage, J.-L., Danchin, E. G. J. & Casane, D. 2007 Where do animal  $\alpha$ -amylases come from? An interkingdom trip. *FEBS Lett.* **581**, 3927–3935. (doi:10.1016/j.febslet.2007.07.019)
- Dagan, T. & Martin, W. 2007 Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc. Natl Acad. Sci. USA* **104**, 870–875. (doi:10.1073/pnas.0606318104)
- Darling, A. E., Miklós, I. & Ragan, M. A. 2008 Selection on genome arrangement in circular bacterial chromosomes. *PLoS Genet.* **4**, e1000128. (doi:10.1371/journal.pgen.1000128)
- Daubin, V. & Ochman, H. 2004 Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res.* **14**, 1036–1042. (doi:10.1101/gr.2231904)
- Daubin, V., Lerat, E. & Perrière, G. 2003 The source of laterally transferred genes in bacterial genomes. *Genome Biol.* **4**, R57. (doi:10.1186/gb-2003-4-9-r57)
- Davids, W. & Zhang, Z. 2008 The impact of horizontal gene transfer in shaping operons and protein interaction networks—direct evidence of preferential attachment. *BMC Evol. Biol.* **8**, 23. (doi:10.1186/1471-2148-8-23)
- De Gelder, L., Williams, J. J., Ponciano, J. M., Sota, M. & Top, E. M. 2008 Adaptive plasmid evolution results in host-range expansion of a broad-host-range plasmid. *Genetics* **178**, 2179–2190. (doi:10.1534/genetics.107.084475)
- DeMars, R. & Weinfurter, J. 2008 Interstrain gene transfer in *Chlamydia trachomatis in vitro*: mechanism and significance. *J. Bacteriol.* **190**, 1605–1614. (doi:10.1128/JB.01592-07)
- Denamur, E. *et al.* 2000 Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* **103**, 711–721. (doi:10.1016/S0092-8674(00)00175-6)
- de Vries, J. & Wackernagel, W. 2005 Microbial horizontal gene transfer and the DNA release from transgenic crop plants. *Plant Soil* **266**, 91–105. (doi:10.1007/s11104-005-4783-x)
- Doolittle, W. F. 2009 The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. *Phil. Trans. R. Soc. B* **364**, 2221–2228. (doi:10.1098/rstb.2009.0032)
- Dorman, C. J. 2007 H-NS, the genome sentinel. *Nat. Rev. Microbiol.* **5**, 157–161. (doi:10.1038/nrmicro1598)
- Drake, J. W. 1991 A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA* **88**, 7160–7164. (doi:10.1073/pnas.88.16.7160)
- Dunning Hotopp, J. C. *et al.* 2007 Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**, 1753–1756.



- Eisenberg, E. & Levanon, E. Y. 2003 Preferential attachment in the protein network evolution. *Phys. Rev. Lett.* **91**, 138701. (doi:10.1103/PhysRevLett.91.138701)
- Eppinger, M., Baar, C., Raddatz, G., Huson, D. H. & Schuster, S. C. 2004 Comparative analysis of four Campylobacteriales. *Nat. Rev. Microbiol.* **2**, 872–885. (doi:10.1038/nrmicro1024)
- Filée, J., Forterre, P. & Laurent, J. 2003 The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Res. Microbiol.* **154**, 237–243. (doi:10.1016/S0923-2508(03)00066-4)
- Fournier, G. P., Huang, J. & Gogarten, J. P. 2009 Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life. *Phil. Trans. R. Soc. B* **364**, 2229–2239. (doi:10.1098/rstb.2009.0033)
- Fraser, J. S., Yu, Z., Maxwell, K. L. & Davidson, A. R. 2006 Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J. Mol. Biol.* **359**, 496–507. (doi:10.1016/j.jmb.2006.03.043)
- Friesen, T. L. *et al.* 2006 Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat. Genet.* **38**, 953–956. (doi:10.1038/ng1839)
- 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. (doi:10.1038/nrmicro1235)
- Galtier, N. 2007 A model of horizontal gene transfer and the bacterial phylogeny problem. *Syst. Biol.* **56**, 633–642. (doi:10.1080/10635150701546231)
- 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. (doi:10.1371/journal.pbio.0030316)
- Gelfand, M. S. & Rodionov, D. A. 2008 Comparative genomics and functional annotation of bacterial transporters. *Phys. Life Rev.* **5**, 22–49. (doi:10.1016/j.plrev.2007.10.003)
- Gerdes, K., Møller-Jensen, J. & Bugge Jensen, R. 2000 Plasmid and chromosome partitioning: surprises from phylogeny. *Mol. Microbiol.* **37**, 455–466. (doi:10.1046/j.1365-2958.2000.01975.x)
- Gillespie, J. J., Beier, M. S., Rahman, M. S., Ammerman, N. C., Shallom, J. M., Purkayastha, A., Sobral, B. S. & Azad, A. F. 2007 Plasmids and rickettsial evolution: insight from *Rickettsia felis*. *PLoS ONE* **2**, e266. (doi:10.1371/journal.pone.0000266)
- Gogarten, J. P. & Townsend, J. P. 2005 Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **3**, 679–687. (doi:10.1038/nrmicro1204)
- Gribaldo, S. & Brochier-Armanet, C. 2006 The origin and evolution of Archaea: a state of the art. *Phil. Trans. R. Soc. B* **361**, 1007–1022. (doi:10.1098/rstb.2006.1841)
- Hao, W. & Golding, G. B. 2006 The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome Res.* **16**, 636–643. (doi:10.1101/gr.4746406)
- Heinemann, J. A. & Sprague, G. F. 1989 Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**, 205–209. (doi:10.1038/340205a0)
- Horn, M. & Wagner, M. 2001 Evidence for additional genus-level diversity of Chlamydiales in the environment. *FEMS Microbiol. Lett.* **204**, 71–74. (doi:10.1111/j.1574-6968.2001.tb10865.x)
- Huang, J. & Gogarten, J. P. 2006 Ancient horizontal gene transfer can benefit phylogenetic reconstruction. *Trends Genet.* **22**, 361–366. (doi:10.1016/j.tig.2006.05.004)
- Inagaki, Y., Susko, E. & Roger, A. J. 2006 Recombination between elongation factor 1 $\alpha$  genes from distantly related archaeal lineages. *Proc. Natl Acad. Sci. USA* **103**, 4528–4533. (doi:10.1073/pnas.0600744103)
- Jain, R., Rivera, M. C. & Lake, J. A. 1999 Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 3801–3806. (doi:10.1073/pnas.96.7.3801)
- Jain, R., Rivera, M. C., Moore, J. E. & Lake, J. A. 2003 Horizontal gene transfer accelerates genomes innovation and evolution. *Mol. Biol. Evol.* **20**, 1598–1602. (doi:10.1093/molbev/msg154)
- Jenkins, C. *et al.* 2002 Genes for the cytoskeletal protein tubulin in the bacterial genus *Prostheco bacter*. *Proc. Natl Acad. Sci. USA* **99**, 17 049–17 054. (doi:10.1073/pnas.012516899)
- Kettler, G. C. *et al.* 2007 Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.* **3**, e231. (doi:10.1371/journal.pgen.0030231)
- Kim, P. M., Korb, J. O. & Gerstein, M. B. 2007 Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. *Proc. Natl Acad. Sci. USA* **104**, 20 274–20 279. (doi:10.1073/pnas.0710183104)
- Konstantinidis, K. T., Ramette, A. & Tiedje, J. M. 2006 The bacterial species definition in the genomic era. *Phil. Trans. R. Soc. B* **361**, 1929–1940. (doi:10.1098/rstb.2006.1920)
- Kunin, V. & Ouzounis, C. A. 2003 The balance of driving forces during genome evolution in prokaryotes. *Genome Res.* **13**, 1589–1594. (doi:10.1101/gr.1092603)
- Kunin, V., Goldovsky, L., Darzentas, N. & Ouzounis, C. A. 2005 The net of life: reconstructing the microbial phylogenetic network. *Genome Res.* **15**, 954–959. (doi:10.1101/gr.3666505)
- Lacroix, B., Tzfira, T., Vainstein, A. & Citovsky, V. 2006 A case of promiscuity: *Agrobacterium*'s endless hunt for new partners. *Trends Genet.* **22**, 29–37. (doi:10.1016/j.tig.2005.10.004)
- Lawrence, J. G. & Ochman, H. 2002 Reconciling the many faces of lateral gene transfer. *Trends Microbiol.* **20**, 1–3.
- Lefebvre, T. & Stanhope, M. J. 2007 Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol.* **8**, R71. (doi:10.1186/gb-2007-8-5-r71)
- Lerat, E., Daubin, V. & Moran, N. A. 2003 From gene trees to organismal phylogeny in prokaryotes: the case of the  $\gamma$ -proteobacteria. *PLoS Biol.* **1**, e19. (doi:10.1371/journal.pbio.0000019)
- Lerat, E., Daubin, V., Ochman, H. & Moran, N. A. 2005 Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* **3**, e130. (doi:10.1371/journal.pbio.0030130)
- Lercher, M. J. & Pál, C. 2008 Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol. Biol. Evol.* **25**, 559–567. (doi:10.1093/molbev/msm283)
- Lima, W. C., Paquola, A. C. M., Varani, A. M., Van Sluys, M.-A. & Menck, C. F. M. 2008 Laterally transferred genomic islands in Xanthomonadales related to pathogenicity and primary metabolism. *FEMS Microbiol. Lett.* **281**, 87–97. (doi:10.1111/j.1574-6968.2008.01083.x)
- Lin, Z., Nei, M. & Ma, H. 2007 The origins and early evolution of DNA mismatch repair genes—multiple horizontal gene transfers and co-evolution. *Nucl. Acids Res.* **35**, 7591–7603. (doi:10.1093/nar/gkm921)
- Lin, C. H., Bourque, G. & Tan, P. 2008 A comparative synteny map of *Burkholderia* species links large-scale genome rearrangements to fine-scale nucleotide variation in prokaryotes. *Mol. Biol. Evol.* **25**, 549–558. (doi:10.1093/molbev/msm282)
- Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F. & Chisholm, S. W. 2004 Photosynthesis genes in *Prochlorococcus* cyanophage. *Proc. Natl Acad. Sci. USA* **101**, 11 013–11 018. (doi:10.1073/pnas.0401526101)

- Makarova, K. *et al.* 2006 Comparative genomics of the lactic acid bacteria. *Proc. Natl Acad. Sci. USA* **103**, 15 611–15 616. (doi:10.1073/pnas.0607117103)
- Martin, W. *et al.* 2002 Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl Acad. Sci. USA* **99**, 12 246–12 251. (doi:10.1073/pnas.182432999)
- Mathee, K. *et al.* 2008 Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc. Natl Acad. Sci. USA* **105**, 3100–3105. (doi:10.1073/pnas.0711982105)
- Mau, B., Glasner, J. D., Darling, A. E. & Perna, N. T. 2006 Genome-wide detection and analysis of homologous recombination among sequenced strains of *Escherichia coli*. *Genome Biol.* **7**, R44. (doi:10.1186/gb-2006-7-5-r44)
- McMillan, D. J., Beiko, R. G., Geffers, R., Buer, J., Schouls, L. M., Vlaminckx, B. J., Wannet, W. J., Sriprakash, K. S. & Chhatwal, G. S. 2006 Genes for the majority of group A streptococcal virulence factors and extracellular surface proteins do not confer an increased propensity to cause invasive disease. *Clin. Infect. Dis.* **43**, 884–891. (doi:10.1086/507537)
- Medini, D., Donati, C., Tettelin, H., Massignani, V. & Rappuoli, R. 2005 The microbial pan-genome. *Curr. Opin. Genet. Devel.* **15**, 589–594. (doi:10.1016/j.gde.2005.09.006)
- Milkman, R. & Bridges, M. M. 1990 Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. *Genetics* **126**, 505–517.
- Mitreva, M., Blaxter, M. L., Bird, D. M. & McCarter, J. P. 2005 Comparative genomics of nematodes. *Trends Genet.* **21**, 573–581. (doi:10.1016/j.tig.2005.08.003)
- Nakamura, Y., Itoh, T., Matsuda, N. & Gojobori, T. 2004 Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* **36**, 760–766. [Corrigendum in *Nat. Genet.* **36**, 1126, 2004.] (doi:10.1038/ng1381)
- Nakashima, K., Yamada, L., Satou, Y., Azuma, J. & Satoh, N. 2004 The evolutionary origin of animal cellulose synthase. *Devel. Genes Evol.* **214**, 81–88. (doi:10.1007/s00427-003-0379-8)
- Navarre, W. W., McClelland, M., Libby, S. J. & Fang, F. C. 2007 Silencing of xenogeneic DNA by H-NS—facilitation of lateral gene transfer in bacteria by a defence system that recognizes foreign DNA. *Genes Devel.* **21**, 1456–1471. (doi:10.1101/gad.1543107)
- Nelson, K. E. *et al.* 1999 Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329.
- Nielsen, K. M., Bones, A. M., Smalla, K. & Elsas, J. D. 1998 Horizontal gene transfer from transgenic plants to terrestrial bacteria—a rare event? *FEMS Microbiol. Rev.* **22**, 79–103.
- Niemeyer, J. & Gessler, F. 2002 Determination of free DNA in soils. *J. Plant Nutr. Soil Sci.* **165**, 121–124. (doi:10.1002/1522-2624(200204)165:2<121::AID-JPLN1111121>3.0.CO;2-X)
- Nystedt, B., Frank, A. C., Tholleson, M. & Andersson, S. G. E. 2008 Diversifying selection and concerted evolution of a type IV secretion system in *Bartonella*. *Mol. Biol. Evol.* **25**, 287–300. (doi:10.1093/molbev/msm252)
- Ogata, H., Renesto, P., Audic, S., Robert, C., Blanc, G., Fournier, P. E., Parinello, H., Claverie, J. M. & Raoult, D. 2005 The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol.* **3**, e248. (doi:10.1371/journal.pbio.0030248)
- Ogata, H., La Scola, B., Audic, S., Renesto, P., Blanc, G., Robert, C., Fournier, P. E., Claverie, J. M. & Raoult, D. 2006 Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.* **2**, e76. (doi:10.1371/journal.pgen.0020076)
- Oikonomou, P. & Cluzel, P. 2006 Effects of topology on network evolution. *Nat. Phys.* **2**, 532–536. (doi:10.1038/nphys359)
- Oliver, R. P. & Solomon, P. S. 2008 Recent fungal diseases of crop plants: is lateral gene transfer a common theme? *Mol. Plant Microbe Interact.* **21**, 287–293. (doi:10.1094/MPMI-21-3-0287)
- Ouzounis, C. A. 2005 Ancestral state reconstructions for genomes. *Curr. Opin. Genet. Dev.* **15**, 595–600. (doi:10.1016/j.gde.2005.09.011)
- Pál, C., Papp, B. & Lercher, M. J. 2005 Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* **37**, 1372–1375. (doi:10.1038/ng1686)
- Papke, R. T., Koenig, J. E., Rodríguez-Valera, F. & Doolittle, W. F. 2004 Frequent recombination in a saltern population of *Halorubrum*. *Science* **306**, 1928–1929.
- Papp, B., Pál, C. & Hurst, L. D. 2003 Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194–197. (doi:10.1038/nature01771)
- Ponting, C. P., Aravind, L., Schultz, J., Bork, P. & Koonin, E. V. 1999 Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* **289**, 729–745. (doi:10.1006/jmbi.1999.2827)
- Poptsova, M. S. & Gogarten, J. P. 2007 The power of phylogenetic approaches to detect horizontally transferred genes. *BMC Evol. Biol.* **7**, 45. (doi:10.1186/1471-2148-7-45)
- Price, M. N., Dehal, P. S. & Arkin, A. P. 2008 Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol.* **9**, R4. (doi:10.1186/gb-2008-9-1-r4)
- Ragan, M. A. 2002 Reconciling the many faces of lateral gene transfer. *Trends Microbiol.* **10**, 4. (doi:10.1016/S0966-842X(01)02283-1)
- 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. (doi:10.1016/j.tim.2005.11.004)
- Raymond, J., Siefert, J. L., Staples, C. R. & Blankenship, R. E. 2004 The natural history of nitrogen fixation. *Mol. Biol. Evol.* **21**, 541–554. (doi:10.1093/molbev/msh047)
- Redfield, R. J. 2001 Do bacteria have sex? *Nat. Rev. Gene* **2**, 634–639. (doi:10.1038/35084593)
- Renesto, P., Ogata, H., Audic, S., Claverie, J.-M. & Raoult, D. 2005 Some lessons from *Rickettsia* genomics. *FEMS Microbiol. Rev.* **29**, 99–117. (doi:10.1016/j.femsre.2004.09.002)
- Richards, T. A., Hirt, R. P., Williams, B. A. P. & Embley, T. M. 2003 Horizontal gene transfer and the evolution of parasitic protozoa. *Protist* **154**, 17–32. (doi:10.1078/143446103764928468)
- Richards, T. A. *et al.* 2006 Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot. Cell* **5**, 1517–1531. (doi:10.1128/EC.00106-06)
- Richardson, A. O. & Palmer, J. D. 2007 Horizontal gene transfer in plants. *J. Exp. Bot.* **58**, 1–9. (doi:10.1093/jxb/erl148)
- Schlüter, A., Szczepanowski, R., Pühler, A. & Top, E. M. 2007 Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. *FEMS Microbiol. Rev.* **31**, 449–477. (doi:10.1111/j.1574-6976.2007.00074.x)
- Schmidt, K. D., Tummeler, B. & Romling, U. 1996 Comparative genome mapping of *Pseudomonas aeruginosa* PAO with *P. aeruginosa* C, which belongs to a major clone

- in cystic fibrosis patients and aquatic habitats. *J. Bacteriol.* **178**, 85–93.
- Schoen, C. *et al.* 2008 Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*. *Proc. Natl Acad. Sci. USA* **105**, 3473–3478. (doi:10.1073/pnas.0800151105)
- Shi, T. & Falkowski, P. G. 2008 Genome evolution in cyanobacteria: the stable core and the variable shell. *Proc. Natl Acad. Sci. USA* **105**, 2510–2515. (doi:10.1073/pnas.0711165105)
- Sonea, S. & Panisset, M. 1983 *A new bacteriology*. Boston, MA: Jones & Bartlett.
- Springael, D. & Top, E. M. 2004 Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. *Trends Microbiol.* **12**, 53–58. (doi:10.1016/j.tim.2003.12.010)
- Sriramulu, D. D. 2008 Adaptive expression of foreign genes in the clonal variants of bacteria: from proteomics to clinical application. *Proteomics* **8**, 882–892. (doi:10.1002/pmic.200700811)
- Stackebrandt, E. & Goebel, B. M. 1994 A place for DNA–DNA reassociation and 16S rDNA sequence analysis in the present species definition in bacteriology. *Int J. Syst. Bacteriol.* **44**, 846–849.
- Staley, J. T. 2006 The bacterial species dilemma and the genomic–phylogenetic species concept. *Phil. Trans. R. Soc. B* **361**, 1899–1909. (doi:10.1098/rstb.2006.1914)
- Suchard, M. 2005 Stochastic models for horizontal gene transfer: taking a random walk through tree space. *Genetics* **170**, 419–431. (doi:10.1534/genetics.103.025692)
- Tettelin, H. *et al.* 2005 Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial ‘pan-genome’. *Proc. Natl Acad. Sci. USA* **102**, 13 950–13 955. (doi:10.1073/pnas.0506758102)
- Thomas, C. M. & Nielsen, K. M. 2005 Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**, 711–721. (doi:10.1038/nrmicro1234)
- Thomason, B. & Read, T. D. 2006 Shuffling bacterial metabolomes. *Genome Biol.* **7**, 204. (doi:10.1186/gb-2006-7-2-204)
- Tobe, T. *et al.* 2006 An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proc. Natl Acad. Sci. USA* **103**, 14 941–14 946. (doi:10.1073/pnas.0604891103)
- Toth, I. K., Pritchard, L. & Birch, P. R. J. 2006 Comparative genomics reveals what makes an enterobacterial plant pathogen. *Annu. Rev. Phytopathol.* **44**, 305–336. (doi:10.1146/annurev.phyto.44.070505.143444)
- van der Does, H. C. & Rep, M. 2007 Virulence genes and the evolution of host specificity in plant-pathogenic fungi. *Mol. Plant Microbe Interact.* **20**, 1175–1182. (doi:10.1094/MPMI-20-10-1175)
- Vernikos, G. S. & Parkhill, J. 2008 Resolving the structural features of genomic islands: a machine learning approach. *Genome Res.* **18**, 331–342. (doi:10.1101/gr.7004508)
- Vlassov, V. V., Laktionov, P. P. & Rykova, E. Y. 2007 Extracellular nucleic acids. *BioEssays* **29**, 654–667. (doi:10.1002/bies.20604)
- von Passel, M. W., Marri, P. R. & Ochman, H. 2008 The emergence and fate of horizontally acquired genes in *Escherichia coli*. *PLoS Comput. Biol.* **4**, e100059.
- Welch, R. A. *et al.* 2002 Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl Acad. Sci.* **99**, 17 020–17 024. (doi:10.1073/pnas.252529799)
- Wellner, A., Lurie, M. N. & Gophna, U. 2007 Complexity, connectivity, and duplicability as barriers to lateral gene transfer. *Genome Biol.* **8**, R156. (doi:10.1186/gb-2007-8-8-r156)
- Woese, C. R. 2000 Interpreting the universal phylogenetic tree. *Proc. Natl Acad. Sci. USA* **97**, 8392–8396. (doi:10.1073/pnas.97.15.8392)
- Yap, H. P., Wang, J. F. & Zhang, Z. F. 1999 Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J. Bacteriol.* **181**, 5201–5209.
- Yčas, M. 1974 On earlier states of the biochemical system. *J. Theor. Biol.* **44**, 145–160. (doi:10.1016/S0022-5193(74)80035-4)
- Zaneveld, J. R., Nemergut, D. R. & Knight, R. 2008 Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiology* **154**, 1–15. (doi:10.1099/mic.0.2007/011833-0)
- Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F. & Papke, R. G. 2006 Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* **16**, 1099–1108. (doi:10.1101/gr.5322306)