trendiness factor is not such an issue. Not to mention the waste of every scientist's time because new rounds of review are needed at every journal. One idea is every journal would have a front end and a back end. Papers would be submitted to the journal and receive a technical review. Then the editorial board would make the decision as to which of these papers would be in the 'front journal' and which in the 'back end' journal. This would be more like a newspaper, where journalists write articles, subeditors check them, and the editors decide where they should go in the newspaper.

You've had conventional success in science: to what do you attribute this? For me, the important things were the mentors that I had during my education. Not only John White and Tim Mitchison, who I have already mentioned, but Eric Karsenti and Kai Simons at EMBL, as well as Nick Crispe and Jim Morgan when I was an undergraduate. These people all took me under their wing at crucial stages in my career, ensuring that I did not fall into the common traps of young scientists, hubris and lack of ambition. I was also lucky to stumble on the field of cell organization just as it was reawakening from its slumber since the 1920s and E.B. Wilson. It is still amazing to think that when I was a PhD student, we did not know of any molecules required for the division of the cell, and now we know most of them.

What next? Well the cataloguing has been a tremendous success, but it has not told us how cells work. The next stage will be to understand how the collective properties of all these molecules give rise to structures that are many orders of magnitude bigger than the molecules themselves. This will involve a component of theory and we are at an exciting stage where physics biologists and mathematicians will have to work together closely to understand these problems. I don't think it will be possible in the future for isolated labs to make major contributions — the skill sets required are too diverse. This suggests teamwork and collaboration — which I can only applaud.

Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), Pfotenhauerstrasse 108, 01307 Dresden, Germany.
E-mail: hyman@mpi-cbg.de

Primer

Lateral gene transfer

Olga Zhaxybayeva[1] and W. Ford Doolittle[2]

The four disparate images shown in Figure 1 have this in common: each represents a radical adaptation that would not have happened had *lateral gene transfer* (LGT), also known as *horizontal* gene transfer (HGT), not been the powerful evolutionary force we now know it to be. Those who study the phenomenon are still struggling to quantitatively assess LGT as a process or processes and accommodate its implications for how patterns in nature should be represented — such as the existence of definable species or a meaningful universal Tree of Life. But all agree that the exchange of genetic information across species lines — which is how we will define LGT in this primer — is far more pervasive and more radical in its consequences than we could have guessed just a decade ago. Both prokaryotes (bacteria and archaea) and eukaryotes have experienced LGT, though its potential as a source of novel adaptations and as a challenge to phylogenetics are so far more obvious and better understood for prokaryotes, as are the mechanisms by which it is effected.

How we detect and measure LGT

The overwhelmingly dominant pattern of heredity is of course 'vertical descent' — the passage of

Figure 1. Four novel adaptations made possible by LGT.
Upper left: a solar saltern in Eliat, Israel. Upper right: Utagawa Hiroshige's *Bowl of Sushi* (detail of woodcut). Lower left: pepper plant roots infected by root-knot nematode. Lower right: pea aphids (*Acyrthosiphon pisum*) exhibiting green, red and yellow color polymorphisms. See text for details. Image credits: Upper left, R. Thane Papke; upper right, Wikimedia Commons; lower left, Scott Bauer, USDA agricultural research service, bugwood.org; lower right, Charles Hedgcock and Nancy Moran.

genes down through generations by normal reproductive and replication processes within species, generally taken to include recombination within sexually reproducing populations. LGT events violate this pattern and can be detected by phylogenetic incongruence, patchy distribution (presence or absence patterns) or compositional anomalies, and best by two or three of these together.

Figure 2 provides an illustration of phylogenetic incongruence from our own work. Several *Prochlorococcus* isolates, though as expected clustering among other cyanobacteria in most gene trees, appear solidly embedded within the gamma-proteobacteria in a tree based on threonyl-tRNA synthetase protein sequences. A single gene replacement by LGT from a gamma-proteobacterium is by far the most parsimonious explanation of this tree. Although technically such a pattern might also result through differential loss of synthetase genes from a common ancestral genome bearing multiple (at least four) threonyl-tRNA synthetase genes, such a genome is unknown today, and very many independent loss events would need to be inferred.

Detecting LGT by patchy distributions (gene presence/absence comparisons) is also parsimony-based. The lower right corner of Figure 1 illustrates such a case. Of all known animals, only two closely-related aphid species have certain genes for carotenoid biosynthesis, genes that are otherwise common among bacteria, archaea, plants and fungi. Explaining this by differential loss would require that the last common ancestor of all animals carried these genes and that there have since been scores of independent gene losses in lineages branching off below the two aphid species. LGT would be by far the favored explanation. This conclusion would be made rock-solid if phylogenetic analysis of the aphid genes linked them to *specific* groups within the phylogeny of fungi, showing incongruence to expectation. This is actually the case: more detail is provided in our penultimate section.

There can also be informative base-compositional or codon usage differences characterizing recent arrivals by LGT, if the gene came from a donor with different base composition or codon usage pattern. And sometimes there will be tell-tale traces of the mechanism of transfer, most obviously when independently replicating and transmissible agents such as plasmids or well-defined mobile genetic elements are the causes of prokaryotic LGT.

It is nevertheless never easy to tell just how many of a genome's genes have arrived by LGT, although estimates are often offered when a new genome sequence is published. Indeed, the question really makes no sense unless a time-frame is specified. For instance, it is often said that there have been no bacterial transfers into the human genome, but without specifying 'since when' — since our divergence from Neanderthals, or chimpanzees and bonobos, or invertebrates, or plants, or indeed from bacteria? If one considers the bacterial contribution to the early formation of eukaryotic genomes, then *most* of our gene families are bacterial transfers! A similarly frustrating imprecision dogs supposedly quantitative estimates of the foreign gene content of prokaryotic genomes, estimates on which strong arguments about the importance or unimportance of LGT have nevertheless been founded.

To be sure, faithful vertical descent is the normal course of events. But even a vanishingly tiny frequency of LGT — one event for every 10^{10} vertical replications, we figure — would be enough to ensure that *no* gene in any modern genome has an unbroken history of vertical descent back to some hypothetical last universal common ancestor. Given that we know (1) that among contemporary prokaryotes gene transfer can be much more frequent than once per 10^{10} vertical replications, (2) that for the first two or more billion years of Life's history the biota was exclusively prokaryotic, and (3) that no individual pairs of genes have strong enough phylogenetic signal to prove that their trees are congruent at such a depth, there is no justification for taking vertical descent as the default scenario. When it comes to inferring ancient LGTs from ambiguous data, absence of evidence is not adequate evidence of absence.

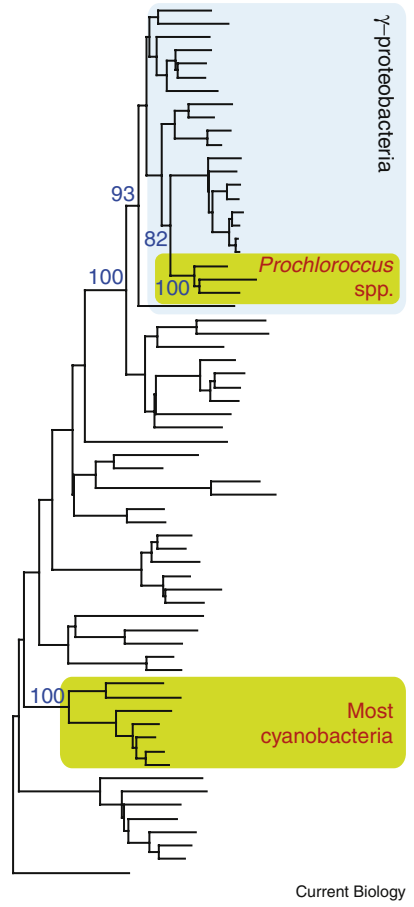

Figure 2. Evidence of LGT through observed phylogenetic incongruence.
The phylogenetic tree of bacterial threonyl tRNA synthetases suggests that some cyanobacteria (*Prochlorococcus* spp.) obtained this gene from gamma-proteobacteria, as they are embedded (with robust statistical support) inside the gamma-proteobacterial clade. The figure is based on Figure 5 in Zhaxybayeva *et al*. (2006).

LGT in prokaryotes: extent, mechanisms, reasons and implications

LGT was foundational to molecular biology, and the three main routes by which bacteria (and as far as is known, archaea) take up 'foreign' DNA have all been known to us for more than fifty years. In the early 1940s, Avery, MacLeod and MacCarty used *transformation* (cellular uptake of 'naked' DNA) to show that DNA is the carrier of genetic information. And bacterial *conjugation* (mating by the formation of cell-to-cell connections) and *transduction* (conveyance of genes from one host to the next by bacteriophages) provided the tools with which exquisitely detailed genetic maps were constructed, long before the day when a whole genome could be sequenced during lunch.

That LGT might be an important mode of adaptation has been widely known for almost as long: since the early 1960s, when the work of Japanese microbiologists, proving that the rise of antibiotic resistance among pathogenic bacteria in hospitals was due to the inter-specific spread of plasmids, first became available in English. But the extent and more general importance of LGT was not obvious until whole genome sequences started to appear, in the last half of the 1990s.

In 1999, about a quarter of the genes in the genome of *Thermotoga maritima*, a hyperthermophilic bacterium, were found more similar to genes in Archaea (the sequenced representatives of which at the time were also hyperthermophiles). Now that there are a dozen more members of the Thermotogales sequenced, the chimeric nature of their genomes is even more amazing. Although the estimated archaeal content has dropped to about ten percent, more than ten times as many assignable genes in their genomes group them with thermophilic clostridia and relatives than with the Aquificales, the group considered on the basis of traditional phylogenetic analyses (using sequences of 16S rRNA and translational proteins) to be their closest relatives.

More typically, a newly sequenced bacterial or archaeal genome will be described as having from a few percent up to a half of its genes as transfers, but such estimates are very method-dependent, often flawed by the false default assumption cited above, and deeply affected by how many and how close are the comparator genomes. Particularly revealing, then, are comparisons among strains of what are considered the same species, where genome-to-genome variation in gene content can be as much as 40%. A recent comparison of the genomes of 61 *Escherichia coli* strains is instructive. Ranging in size from 4,157 to 5,315 gene families, together these genomes share only 933 gene families, defining an *E. coli* 'core' genome. Making up the difference in each genome would be more than three thousand 'accessory' gene families, present in only some (as few as one) of the 61 strains. Each newly sequenced strain reveals more accessory gene families and these plus the core collectively make up what is called *E. coli*'s 'pan-genome'. This now stands at 15,741 gene families — and still growing. Assuming, as seems reasonable, that the ancestral *E. coli* had a typical sized genome, about two-thirds of the current pan-genome must owe its existence to LGT from *other species*.

Who benefits?

Certainly many individual transfers can be seen as adaptations (Figure 1), but that does not necessarily rationalize the existence of the mechanisms that made them possible. Transformation as a mechanism of DNA uptake, for instance, is under control of genes in the recipient and must make sense in terms of advantages to those genes. Assuming adaptive evolution through recombination to be the sole benefit is problematic in the same way as is assuming this to be the sole benefit of sex. So why so many bacterial species boast evolved multi-gene transformations systems is still a mystery. LGT *via* transduction or conjugation may on the other hand be understood as directly or indirectly promoting the spread of those genes of the bacteriophages or conjugative elements (such as many plasmids) that cause it, and hence 'selfish'.

An example of direct benefit would be genes for components of the photosynthetic apparatus carried by bacteriophage infecting cyanobacteria. Although these will allow recombination of such genes across species lines, their real 'purpose' is presumably to drive infected cells to crank out just that many more virus progeny before expiring. An example of indirect benefit, in which genetic systems effecting transfer prosper through helping their hosts, would be the antibiotic resistance determinants that first called attention to LGT. The global antibiotic 'resistome', the collection of genes that confer some resistance to our increasingly ineffective armoury of antimicrobials, is now known to be enormous and distributed broadly among non-pathogens in the environment. What we have encouraged our pathogens to do is sample it by LGT, and there is evidence that they continue to get better at doing this, in terms of the activity and number of encoded resistance determinants and mobility of their genetic carriers.

Bacteria and archaea house a vast menagerie of such potential LGT carriers, or mobile genetic elements, ranging from simple insertion sequences less than a kilobase long and bearing little but a gene for a transposase, to conjugative transposons or integrated conjugative elements that can be several hundred kilobases long and carry genes for their own integration and excision into and from host chromosomes, virulence and drug resistance determinants, catabolism of bizarre substrates, or symbiotic association with eukaryotic hosts. Sometimes such elements bear genes that sense when conjugative transfer might be a good idea (the presence of antibiotic or other stressors, dense populations of potential recipients) and set the process in train.

More generally, prokaryotic genomes seem to comprise: (1) relatively more conserved and generally (over the short-term) syntenic regions — in which core genes and those of broader distribution and better known function are preferentially but not exclusively found — interspersed with (2) variously well defined *genomic islands* which may or may not themselves function as mobile genetic elements of the above sorts. Genomic islands are also preferred resting places for later-arriving mobile genetic elements, and sites of much gene gain and loss: it is not surprising that variable accessory genes distinguishing *E. coli* strains are differentially found in islands embedded within a relatively more stable genomic framework. It is likely that all prokaryotic genomes harbor an island or two. Islands in the numerically dominant marine cyanobacterium *Prochlorococcus* were first identified by comparing sequenced genomes to random reads from a metagenomic library of oceanic DNA. Islands, because of their highly variable gene content, comprise sequences that are less well represented in such a library.

It is popular to consider that the pangenomes of all prokaryotic species collectively make up a vast meta-metagenomic pool of sequences that can be variously sampled for inclusion in prokaryotic chromosomes, most often through

the agency of phages and plasmids. Although it is hard to see how any system whose function is simply to sample this enormous resource could arise, or how mobile genetic elements or cellular lineages could themselves benefit directly by helping to enlarge and diversify it, at least two sorts of prokaryotic element seem to make sense only in such contexts. Integrons, comprising an integrase gene and up to 200 short open reading frames (*gene cassettes*) that they have strung together adjacently by site-specific recombination, appear to sample a large cassette metagenome shared within and between species (vibrios most notably) for mutual benefit. And gene transfer agents, phage-like particles encoded by chromosomal genes that do not ensure their own packaging, seem to exist for the sole purpose of contributing to the global gene pool. There are unsolved mysteries here too.

LGT and the Tree of (Prokaryotic) Life

The core of genes shared by all *E. coli* strains is fewer than 1,000. The core shared by all bacteria has been generously estimated at 250 and all prokaryotes collectively share fewer than 100. These genes are, for many reasons, less likely to be exchanged by LGT, not least because many of them are essential, so that LGT must entail replacement through homologous recombination or tightly coupled gain and loss events.

What evidence is there for congruence of these several scores of genes, and to what phylogenetic depth? Indeed, the majority of individual core gene trees do separate the two prokaryotic domains, Archaea and Bacteria, although phylogenetic signals are in general too weak to show within-domain congruence and the branching orders of major subdivisions ('phyla') remain unresolved. The core is small, however, and its tree has been called the 'Tree of One Percent': what should we take it to represent?

Many would argue that the core tracks a unique 'Tree of Cells' — that pattern of successive cell divisions traceable back to one ancient prokaryotic cell, the so-called last universal common ancestor. Support for this might be taken from the fact that various ways of using collectively the sequences of genes that are found only in some genomes – or the patterns of presence and absence of such genes – to construct trees can give tree topologies not too dissimilar to the core, or to the three-domain tree first proposed on the basis of ribosomal RNA sequence information alone. But the same data support the notion that overall prokaryotic genome evolution has been net-like rather than tree-like, and that much of the last universal common ancestor's genome has been replaced by LGT from contemporaneous lineages that would not be represented in any Tree of Cells.

In global analyses, preferential transfers among specific groups start to emerge, forming so-called 'highways of gene sharing'. Many of these highways connect various bacterial and/or archaeal phyla. Some phyla emerge as clearly separated clusters, while for others, extensive LGTs are observed. Depending on its frequency, there is a possibility that LGT is so rampant that the observed relationships may reflect shared histories of transfers and not vertical inheritance. The literature is in conflict on this possibility.

Eukaryotes: the iceberg's tip

In their 2008 review of LGT into eukaryotic genomes (nuclear and organellar), Keeling and Palmer reported that the list of believable cases was already too long to present: we are three years further down that road. Patchy genome sampling and biased detection methods will inevitably over- or under-estimate the phenomenon, but some generalizations seem safe. The sequestration of germ lines, as occurs in animals, is surely a barrier to the evolutionary fixation of transfers. The most willing LGT recipients are protists, especially those that eat bacteria or harbor them as symbionts. Their nuclear genomes will be *repeatedly* exposed to genes that may supplement or irreversibly replace resident copies. Cohabitation with potential donors, anaerobic protists together with anaerobic prokaryotes in the rumen, for instance, will encourage LGT. The genomes of the insect hosts of the bacterium *Wolbachia* have picked up, mostly as non-functional fragments, all of this parasites' genome. There are also some astonishing cases of eukaryote-to-eukaryote LGT, a phenomenon that is surely under-reported because of the sparse sequencing of this domain. Parasitic plants such as *Amborella* and *Striga* have contributed both mitochondrial and nuclear genes to their hosts, the former in high numbers.

Mechanisms of eukaryotic LGT are poorly known. Transfer of genes from food bacteria, symbionts and organelles might be experimentally modeled by the well-documented formation of *numts* and *nupts*, fragments of mitochondrial and plastid DNA that have been taken up by nuclear genomes, as now revealed by many eukaryotic genome sequencing projects. Broad host range LTR retrotransposons or DNA transposons such as *Ty3* or *mariner*, though known so far only to transfer their own genomes interspecifically, clearly do play a role in eukaryote genome evolution.

Keeling and Palmer (2008) suggest that, in spite of the growing list of eukaryotic LGT examples, "there is no reason to think that [LGT] is so prevalent as to undermine efforts to reconstruct a dichotomously branching tree of eukaryote phylogeny, much less call for the replacement of the tree metaphor with a 'web of life' metaphor, as some have controversially suggested for prokaryotes". Indeed, but if, as most would agree, Life's history is predominantly prokaryotic, what then of the combined 'universal' tree?

Salt, sushi, worms and aphids: four remarkable adaptations

Our Figure 1 is meant to highlight four recent publications in which LGT has produced a striking adaptation that simply could not have arisen though mutation, recombination or selection, and therefore falls outside traditional neodarwinian explanation. Many more examples readily come to mind, and of course the vast majority of lateral transfers will not be revealed as such without fine-scale comparative genomic work. Transfers between all three domains are known, though those from eukaryotes to bacteria or archaea will be severely limited by introns, and those between prokaryotic domains might be discouraged by incompatibilities in gene expression machinery.

The upper left photograph, of a saltern in Israel, represents an archaea-to-bacteria transfer, and suggests why different domains of life co-existing in the same environment might exhibit similar phenotypes. The bacterium *Salinibacter ruber* (from the Bacteroides/Chlorobi phylum) and haloarchaea inhabit environments with high (even saturated) salt concentrations. Recent analysis of a *Salinibacter* genome revealed a cluster of genes likely involved in sodium, potassium or chloride and cationic amino acid transport, and genes for transposases possibly involved in the assembly of this 'hypersalinity island' from various bacterial and archaeal sources. The *Salinibacter* genome also encodes four rhodopsins, retinal-bearing proteins using light to pump protons or cations, and to seek or avoid light, essential for life in the salterns. A phylogenetic analysis shows that three out of four rhodopsins present in the genome group with haloarcheal counterparts and their sequences provide additional evidence for functions similar to their haloarchaeal homologs. Broader surveys of rhodopsin gene distribution have shown that rhodopsin genes are exemplars of environmentally-significant capacities transferred frequently across the three domains of life.

The Hiroshige woodcut in the upper right corner of Figure 1 represents a bacterium-to-bacterium LGT that affects our own species, broadening the metabolic capacity of microbial residents of the human gut. Some gut microbiota aid digestion of carbohydrates from plants we eat, through polysaccharide digestive enzymes that the human genome lacks. Some of these enzymes are present only in the gut bacterium *Bacteroides plebeius* and marine bacteria, the latter using them to degrade red algal polysaccharides. Curiously, based on further analysis of metagenomic datasets of human feces, the genes are present in Japanese individuals and absent in North Americans. It is postulated that the genes came to *B. plebeius* from marine bacteria. The hypothesized link that provided contact between marine bacteria and gut microbiota is dietary seaweed, used raw in sushi.

As noted, LGT plays an important and well-documented role in pathogenicity of bacteria. The image at the lower left of Figure 1 shows a case in which bacteria-to-animal LGT has played a role in pathogenicity of eukaryotes as well. The nematode *Meloidogyne incognita* parasitizes plants and its genome contains over 60 transcriptionally active genes for degradation and softening of plant cell walls. Analysis of the evolutionary histories of these gene families reveals that they originated through independent LGT events from different bacterial sources and, once acquired, further underwent expansion through duplication. Among contemporary relatives of gene donors are plant pathogenic soil bacteria and parasitic bacteria affecting the same host plants.

Our final illustration, the lower right image in Figure 1, represents the recently discovered eukaryote-to-eukaryote (fungus-to-animal) transfer that was highlighted earlier in this essay. Moran and Jarvik (2010) report that genes responsible for pea aphid color derive from an ancient gene transfer from fungi. This is the first reported case of the presence in an animal genome of a cluster of carotenoid genes. The transfer event, followed by deletions and mutations, is responsible for the presence of red, green and yellow colorations in the aphids. Red and green polymorphisms have been known to co-occur in natural populations and are both maintained due to the existence of parasites and predators that recognize specifically either one or the other color.

Final thoughts

LGT can introduce radically new phenotypes that mutation and selection might never achieve. Quantitative estimates of its frequency, even if accurate, may underestimate its importance in adaptation and speciation. Opponents of evolution have cheered its challenge to the 'Tree of Life', as if the literal truth of that *simile* (as Darwin called it) were essential for the modern theory of evolution. We, however, take this theory to be simply that understandable ecological and genetic processes, operating over evolutionary time, are adequate to explain existing biological adaptation and diversity. In fact, the ability of LGT to speed complex and radical adaptation makes it even easier to imagine how "from so simple a beginning, endless forms most beautiful and most wonderful have been, and are being, evolved".

Acknowledgments

Authors' work has been supported by the Canadian Institutes for Health Research and the Tula Foundation. O.Z. is also supported by start up funds from West Virginia University.

Further reading

Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., Delong, E.F., and Chisholm, S.W. (2006). Genomic islands and the ecology of *Prochlorococcus*. Science *311*, 1768–1770.

Danchin, E.G., Rosso, M.N., Vieira, P., de Almeida-Engler, J., Coutinho, P.M., Henrissat, B., and Abad, P. (2010). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. Proc. Natl. Acad. Sci. USA *107*, 17651–17656.

Dunning Hotopp, J.C., Clark, M.E., Oliveira, D.C., Foster, J.M., Fischer, P., Muñoz Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., *et al.* (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science *317*, 1753–1756.

Hehemann, J.H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature *464*, 908–912.

Keeling, P.J., and Palmer, J.F. (2008). Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. *9*, 605–618.

Lapierre, P., and Gogarten, J.P. (2009). Estimating the size of the bacterial pan-genome. Trends Genet. *25*, 107–110.

Levin, B.R., and Cornejo, O.E. (2009). The population and evolutionary dynamics of homologous gene recombination in bacterial populations. PLoS Genet. *5*, e1000601.

Lukjancenko, O., Wassenaar, T.M., and Ussery, D.W. (2010). Comparison of 61 sequenced *Escherichia coli* genomes. Microb. Ecol. *60*, 708–720.

Mongodin, E.F., Nelson, K.E., Daugherty, S., Deboy, R.T., Wister, J., Khouri, H., Weidman, J., Walsh, D.A., Papke, R.T., Sanchez Perez, G., *et al.* (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. Proc. Natl. Acad. Sci. USA *102*, 18147–18152.

Moran, N.A., and Jarvik, T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science *328*, 624–627.

Puigbo, P., Wolf, Y.I. and Koonin, E,V. (2010). The tree and net components of prokaryote evolution. Genome Biol. Evol. *2*, 745–756.

Zhaxybayeva, O., Gogarten, J.P., Charlebois, R.L., Doolittle, W.F., and Papke, R.T. (2006). Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. Genome Res. *16*, 1099–1108.

Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., Nelson, K.E., Nesbø, C.L., Doolittle, W.F., Gogarten, J.P., and Noll, K.M. (2009). On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc. Natl. Acad. Sci. USA. *106*, 5865–5870.

[1]Department of Biology, West Virginia University, PO Box 6057, Morgantown, WV 26506, USA. [2]Department of Biochemistry and Molecular Biology, Dalhousie University, 5850 College Street, Halifax, Nova Scotia, Canada, B3H 1X5.
E-mail: ford@dal.ca