Is Horizontal Gene/Domain Transfer important in evolution? How does it affect phylogenetic analysis?

Jennifer Kee

0243336
Is horizontal gene/domain transfer important in evolution? How does it affect phylogenetic analysis?

For several years evolutionary biologists have been debating a controversial issue: the horizontal exchange of genetic material between distantly related species, the extent to which it occurs and its possible adaptive significance. What is commonly accepted is that Horizontal Transfer (HT) is an important evolutionary force in prokaryotes. This is a very broad topic and hence the discussion will be limited to specific debates. The debate as to the influence of horizontal transfer in the early evolution of life will be discussed. Tied in with this topic domain transfer, which is an important evolutionary force particularly in eukaryotes, will be discussed briefly in terms of its occurrence between the three main cell types. The importance of horizontal transmission between unicellular and multicellular organisms and its influence on tree topology will also be considered, using the transfer from bacteria to vertebrates as an example.

Firstly it is important to understand some of the main forces, which are thought to govern the evolution of organisms.

What is evolution and how does it occur?

In biology evolution refers to the processes that have transformed life on Earth from its earliest forms to the vast diversity that characterises it today. Charles Darwin published On the Origin of Species by Means of Natural Selection on November 24th, 1859. Darwin made two main points in his book; first he argued that species were not specially created in their present forms but had evolved from ancestral species. Second he proposed a mechanism for evolution that he termed natural selection. According to the concept of natural selection a population of organisms can change over time as a result of individuals with certain heritable traits leaving more offspring than others. Using the principles of gradualism small changes over a long period of time can cause substantial change overall. Darwin began to perceive that the origins of species and adaptation were closely related processes and anticipated that explaining how such adaptations arise is essential to understanding evolution.

“Descent with modification” is a Darwinian term that implies that all organisms are related through descent from some unknown prototype that lived in the remote past. As descendents of the inaugural organism spilled into various habitats over millions of years they accumulated diverse modifications that fit them to specific ways of life.

Evolution at the population level can be described as a generation-to-generation change in a population’s frequencies of alleles or genotypes, in other words a change in the genetic structure of a population. In what ways can these changes occur? In order to illustrate how changes of this sort may occur the generation of genetic variation in bacteria will be discussed.

The major component of the bacterial genome is a single-stranded DNA molecule arranged in a circle, which is often referred to as the bacterial chromosome. In addition to this chromosome a bacterial cell may possess plasmids, which are smaller circles of DNA. Each plasmid has only a small number of genes the maximum number being about 24. Bacterial cells reproduce by binary fission and this is preceded by replication of the chromosome. Since fission is an asexual process most of the offspring produced by this vertical transmission are genetically identical to the parent with the exception of those in which a mutation has occurred.

Horizontal transfer refers to the transfer of genetic material from one organism to another by processes similar to infection. In addition to mutations genetic recombination generates diversity within bacterial populations. The bacterial mechanisms of genetic recombination and possible sources of HT are transformation, transduction and conjugation. These mechanisms are described in figures 1 and 2.
There are two types of transduction, see the cartoon opposite below.

In generalised transduction random pieces of the host chromosome are packaged within a phage capsid. In specialised transduction a prophage exits the chromosome in such a way that it carries adjacent bacterial genes along with it.

The transfer of DNA by phage requires the donor and the recipient cell to share cell surface receptors for phage binding. Therefore this method of transfer is usually limited to closely related bacteria.

**Transformation**

This is the uptake of naked, foreign DNA from the surrounding environment and the subsequent alteration of a bacterial cells genome. This is a common mode of HT that can mediate the exchange of any part of a chromosome and is most common in bacteria which are naturally transformable. Typically only short fragments of DNA are exchanged.

The figure above is adapted from Campbell 4th Edition.
**Figure 2 describes the processes of Conjugation and Transposition and how they can be related.**

**Conjugation**

This is the direct transfer of genetic material between two bacterial cells that are temporarily joined see the colorized TEM below. The transfer is mediated by conjugal plasmids or conjugal transposons, which are often called the vehicles for genetic exchange.

The spread of antibiotic resistance drew attention to this type of exchange decades ago. Plasmids containing genes for antibiotic resistance can be transferred between bacteria when they are joined. What makes the spread of antibiotic resistance an even greater problem is that a single plasmid may carry as many as ten resistance genes to ten different antibiotics. How can so many different resistance genes accumulate on a single plasmid? The answer involves the other type of mobile genetic element – the transposon. Transposons are pieces of DNA that can move from one location to another in the genome of a cell and are often called “jumping genes”\(^6,8\). This type of genetic element is seen in higher organisms as well as bacteria. Some regions of DNA are more likely to receive a transposon than others but they do not have a single likely target in the genome. This makes transposition a fundamental process, that is less limited than other methods\(^7\), in “genetic shuffling”. The simplest type of transposon is called an insertion sequence. The gene in an insertion sequence codes for the enzyme that catalyses transposition. Non-coding inverted sequences bracket the transposase enzyme sequence. A complex transposon has other genes that essentially “go along for the ride” such as genes for antibiotic resistance. A simplified cartoon of an insertion event is shown below. In the cartoon below part (a) the arrows represent the point at which transposase cuts the DNA. Parts (c) and (d) of the cartoon show the insertion of the transposon. DNA polymerase fills the gaps shown in part (c) and this results in identical DNA segments, or direct repeats, on either side of the transposon. This provides the answer to the question posed earlier. In a bacterial cell a transposon can move into and out of plasmids hence different genes for antibiotic resistance can be inserted into the same plasmid. Equally by the same mechanism genes contained in a complex transposon can be horizontally transferred to another cell via a plasmid.

Insertion sequences account for 1.5% of the E.coli genome in particular. They transpose and cause mutations about once in every ten million generations\(^4\). This is about the same as the mutation rate due to extrinsic causes. Given the rapid proliferation of bacteria transposition of DNA probably plays a significant role in bacterial evolution.

The TEM and the cartoon shown above were adapted from Campbell 4th Edition\(^4\).
Most bacteria have a generation time in the range of 1-3 hours. Some species can double every 20 minutes in an optimal environment. If the latter reproductive rate is sustained a single cell would could give rise to a colony weighing 1 million kg in just 24 hours\(^4\). In most natural environments bacteria compete for space and nutrients. A genetic variation, generated by recombination or by mutation, which conferred a selective advantage on a host cell, could result in the generation of a large number of offspring bearing the same genetic makeup in a relatively short space of time. Adaptive evolution is hence a blend of chance and sorting. Chance in origin of genetic variation by mutation and recombination, and sorting in the workings of selection as it favours the propagation of some chance variations over others. Hence the ability to generate genetic variation is essential for the survival of the organism in question. It must be borne in mind however that not all genetic variations will increase the reproductive success of the individual concerned; neutral mutations are an example.

Each of the mechanisms of genetic exchange described in figures 1 and 2, can introduce sequences of DNA that are unhomologous with the existing DNA of the recipient bacterial cell\(^5\). For example in bacteria genomic islands with different G+C content with respect to the neighbouring regions are generally the result of horizontal gene transfer and recombination events\(^3\). Horizontal transfer events are most definitely important in terms of bacterial evolution.

**Phylogenetic analysis in the study of evolution?**

Organisms are often placed in groups according to their evolutionary history a practice known as classification. To Darwin the natural hierarchy of the Linnaean classification scheme reflected the branching genealogy of the tree of life. The evolutionary history of a species or a group of related species is called phylogeny\(^1\). These genealogies are often described by phylogenetic trees, which trace putative evolutionary relationships.

Many different types of character can be used in phylogenetic analysis\(^1\)\(^9\) and traditionally morphological approaches were used\(^4\) but nucleic acids and proteins are the now most popular analysis methods because they are common to all life forms; this allows both closely and distantly related taxa to be studied. Whole genome analysis can be carried out as well as analysis of a specific gene in different organisms. This all seems very straightforward however there are problems with inferring phylogenies from sequence data and these will be discussed later. The outcome of a phylogenetic study is generally a phylogenetic tree. A tree is a simple way to show evolutionary relationships and it may be rooted or unrooted. A simple example of a phylogenetic tree is shown in figure 3. This is a rooted phylogenetic tree where the root represents a point of divergence of a common ancestor\(^2\); unrooted trees are more common however.

It is beyond the scope of this essay to describe in detail how a phylogenetic tree is constructed. It should suffice to say that in the simplest situation an initial alignment is created, such as a DNA alignment as in this example, then a distance matrix is constructed from which the tree is generated using computer software. If the tree in figure 3 is the “true tree” then it is not difficult to comprehend that human and chimp DNA should be more similar to each other than to gorilla DNA. However if DNA was exchanged horizontally between the gorilla and the chimp for example then this may distort the true tree. The gorilla and the chimp might now be clustered together; this new tree may not represent descent with modification as it actually occurred.
**Figure 3** shows a simple phylogenetic tree indicating the putative evolutionary relationships between gorillas, humans and chimps. The tree was constructed based on DNA analysis.

The tree shown above is an example of a very simple cladogram, or bifurcating tree. The furthermost left vertical line represents a putative last common ancestor from which gorillas, chimpanzees, and humans could have evolved. If the possibility exists that extensive swapping of genes may have occurred between prokaryotic organisms and indeed between higher organisms then it stands to reason that comparisons of the genomes of individual organisms may not indicate how closely related they are by common ancestry. Some researchers are even questioning the existence of a last common ancestor as predicted by Darwinian theory\(^23,24\).

**Is it possible that gene swapping muddles the history of all organisms?**

In the 1960s what was called a Universal phylogenetic tree was constructed in an attempt to unite all the kingdoms into a single phylogenetic empire. Ribosomal RNA was central to this endeavor as not only is the molecule ubiquitous, it exhibits functional constancy, changes slowly in sequence and is experimentally very tractable\(^19\). A very simplified rRNA-based phylogenetic tree is shown figure 4. In the 1990s with the development of genomic studies the opportunity arose to develop this universal tree. Quite the opposite in fact happened; researchers discovered that the phylogenetic trees based on genomic data quite often did not agree with the universal tree. In particular individual gene trees showed incongruency with the RNA tree. This was dubbed the “gene-tree species-tree problem”\(^23\). The initial reaction in the scientific community was one of the “sky falling”\(^23\) and some researchers began to question whether a universal phylogenetic tree could be constructed with any confidence, and in particular if a last common ancestor could ever be divulged, if cross-species HT had occurred pervasively throughout evolution\(^15\).
The Last Common Ancestor debate

On Earth today there exists three distinct cellular designs: the bacterial, the archaeal and the eukaryotic as depicted by figure 4, these are often called the three domains which is a taxonomic rank higher than kingdom\(^4\). The universal phylogenetic tree takes us back to an era before the branching shown in figure 4 occurred. The earth's crust began to solidify 4 billion years ago and the simplest forms of life appeared approximately 3.5 billion years ago. These initial cells were thought to be simpler in design than cells today in the sense that they had less complex, less integrated more modular componentry. It is therefore thought that horizontal gene transfer would have occurred more readily at this time and would have been a more dominant evolutionary force than it is now\(^2\). Life in this form could be thought of as a single diverse gene pool and thus at such a stage evolution could have been in effect communal. If this was the case how then could the different lineages shown in figure 4 evolve?

Woese offered an opinion in as to how this may have happened\(^2\). Woese considers the possibility that in these early evolutionary stages eventually some cellular entities became complex enough to produce unique cell designs. In other words from the communal ancestor new gene pools arose within which HT events continued to occur. It is thought that there was a point where cellular components became sufficiently complex and sufficiently integrated into the cellular fabric that horizontal transfer events had less of an influence. At this point cellular lineages and true organismal lineages arose. Woese suggests that the initial bifurcation in figure 4 marks this point.

What happened after this bifurcation? Did HT occur between the three cellular lineages? In 2002 Syvanen produced a paper\(^2\) suggesting that the genetic code was not fully formed before the divergence of life into the three cell types. The article suggests that arginine and tryptophan evolved after the diversification of the three domains and were spread by HT. Evidence for this hypothesis is based on data suggesting that the enzymes which synthesise arginine and tryptophan have shorter divergence times than the underlying lineages. If this is the case then horizontal transfer played a very significant role in early evolution.

In the same paper Syvanen questions the need, or indeed the ability, to postulate the existence of a last common ancestor, as described by Woese\(^3\) and Doolittle\(^3\). Syvanen argues that the deepest branches of the universal tree can never be distinguished due to extensive HT at these early stages in the evolution of life. The author favours the theory that all of life evolved from multiple lineages.

Comparative genome analysis has revealed major lateral transfer between the three primary cell types\(^5\). However proving HGT in each individual case is difficult. Yanai et al\(^1\) performed a study in which they investigated the possibility that the evolutionary phenomena of horizontal domain transfer and domain fusion may occur between the three cellular lineages. Proteins often form multidomain architectures\(^1, 18, 38\). Each domain is defined as a functional protein in its own right but such domains often comes together by fusion. Proteins that form links by domain fusions are often the components of complex pathways in cells. To link the genes for the various domains often offers a selective advantage as it improves the response of the pathway to various cellular signals. Phylogenetic analyses carried out by Yanai et al suggested that domain exchange followed by domain fusion does occur between the three cellular lineages but it is a relatively rare phenomenon.

Kyrpides\(^2\) disputes the cross-domain significance of HT events. Elsen\(^4\) sums up the concerns of many researchers and writes “the determination and analysis of complete genome sequences has led to the suggestion that horizontal gene transfer may be more extensive than previously appreciated. Many of these studies, however, rely on evidence that could be generated by forces other than gene transfer including selection, variable evolutionary rates, and biased sampling”.
What is the evidence for Horizontal Transfer events in multicellular organisms and in particular between unicellular and multicellular organisms?

The “grand-daddies” of horizontal transfers are believed to have led to the possession by some cells of mitochondria and chloroplasts\textsuperscript{11} see figure 5. It seems that these organelles originated as bacteria\textsuperscript{5,11} and even retain a few of their original genes in mitochondrial and chloroplast DNA. It is thought that perhaps a single bacterium was engulfed by a primitive ancestor and then formed an endosymbiotic relationship with its captor. It is also thought that at some point some of the DNA of the ‘invader’ became part of the ‘host’ nuclear DNA. These are long-standing ideas but it is only recently that it has been possible to positively identify some of the mitochondrial and chloroplast genes as of bacterial origin. If these events did occur, as stated above, then it is obvious that these horizontal gene transfer events had a significant impact on shaping the diversity of life that is seen today.

In 1984 Michael Syvanen published work\textsuperscript{13} that compared the nucleotide sequences from the coding regions of four different species of mammalian B-globins. Syvanen observed that certain regions of the globins had accumulated substitutions in a random fashion as predicted by the molecular clock hypothesis\textsuperscript{14}. Some regions however appeared to have evolved at a much slower rate. Conservation of these regions was hypothesised to be due to what the author called “cross-species gene exchange”. This exchange was thought to have occurred between the mammalian lineages at a time significantly after their evolutionary divergence. The author writes “I postulate that the different germ lines are occasionally exposed to B-globin sequences from other mammalian species (perhaps by viral infection) and that the information on these sequences becomes incorporated by a gene conversion event.” The possible infective viral intermediate is undisclosed in this discussion. In the same article many other examples are cited of possible HT events between different species and in particular between higher organisms.

Finally according to a study commissioned by the British governments Ministry of Agriculture, Fisheries and Food a transposable element called the mariner element has crossed the species barrier seven times in evolutionary history\textsuperscript{12}; at one time between tsetse flies and humans. Andy
Brass and two colleagues at the University of Manchester compared the DNA of 80,000 different organisms using over 5 million sequences. The investigators thought that the transfer between the tsetse fly and humans occurred recently in evolutionary terms but it is not clear in what direction it occurred.

Figure 5 shows a hypothetical phylogeny representing the origins of eukaryotic diversity.

In particular the diagram above shows the postulated horizontal transfer of chloroplasts and mitochondria. These events could be thought of as the “grand daddies” of horizontal gene transfer.

The figure above is adapted from Campbell 4th Edition.
The Human Genome Project produced some surprising findings

A horizontal transfer event can only become fixated in the genome of a multicellular organism if it is transmitted via the germ line. If an individual horizontal transfer event occurred by transposition, or was virally mediated, the odds are against this event occurring in the genome of a multicellular organism. It should also be taken into consideration that not all changes in genetic makeup will have an effect on reproductive success. To estimate the importance of HT events in the multicellular organisms is therefore a difficult matter. There is no general rule as to how much genetic change is required for speciation for example. What can be said is that any processes, such as HT, which can generate variation, have the potential to cause change under the scrutiny of natural selection. For the reasons outlined above horizontal transfers are unlikely to play as much of a role in the evolution of multicellular organisms as in unicellular organisms such as bacteria. Equally exchange between a unicellular organism and a multicellular organism leading to transmission to offspring of the multicellular organism is likely to be a relatively rare event. It therefore surprised many researchers when one of the main conclusions of the International Sequencing Consortium published in 2001 was that 113 genes had been transferred from bacteria to vertebrates.

Various groups set about finding out if these genes represented examples of HT from bacteria to the vertebrate lineage or were they present in both prokaryotes and early eukaryotes but were subsequently lost from all non-vertebrate eukaryotic lineages? Andersson et al asked, “Are there bugs in our genome?”.

Stanhope et al carried out a study in which they phylogenetically analysed 28 proposed HT genes. The results from this study indicated that most of the putative HT genes are present in more anciently derived eukaryotes and can be explained in term of descent through common ancestry. The authors of this article suggest that the reason the initial report did not come to this conclusion was that BLAST searches were used to find the genes in other species and also that non-vertebrate eukaryotes were often overlooked as the source of genes. This was supported by work carried out by Roelofs et al who found that at least 11 of the 113 listed genes had homologues in the primitive eukaryote Dictyostelium.

Phylogenetic principles for rejecting or accepting HT

The branching arrangement of the phylogenetic tree produced from a phylogenetic study is a critical component of a HT assessment. For example if in a phylogenetic tree a vertebrate sequence appears in a cluster of bacterial sequences then the vertebrate gene could have been horizontally transferred from bacteria. Conversely if all eukaryotic sequences are clustered together then HT has probably not occurred. Figure 6 displays phylogenetic trees, which both support and reject HT events between bacteria and vertebrates. All but one of the 28 candidate sequences analysed by Stanhope et al fell into this latter category.

Tanita Casci speculated that the consortium results were not similar to the findings above for several reasons. Firstly that perhaps non-vertebrate orthologues were not actively searched for or were not found. Secondly perhaps too few taxa were sampled and lastly that those taxa that were analysed had been affected by gene loss. Casci writes, “As things stand, the exciting possibility that a bacteria-to-vertebrate HGT event has occurred seems to be fading fast.”
The three trees above are taken from a study carried out by Stanhope et al\textsuperscript{28}. The first tree shows a phylogenetic history in support of bacteria to vertebrate HT. The second tree shows a history that would support transfer in the opposite direction, from vertebrates to bacteria. The third tree rejects any sort of HT event between bacteria and vertebrates. The paralogue sequence represents an ancient non-vertebrate eukaryote. Note that in trees (a) and (b) vertebrates and bacteria are grouped closely together in the tree indicating potential HT events. There is no evidence of this in tree (c) as vertebrate sequences are separated from bacterial sequences by a non-vertebrate eukaryote sequence.
Conclusion

It is clear that horizontal transfer is an important evolutionary force in prokaryotes. However due to the germ line transmission requirement it is unlikely that HT is as important, in terms of evolution, in multicellular organisms.

Regarding the use of phylogenetic trees, in the study of evolution, it is truly proving a difficult task to reconstruct the history of life on earth. The Universal Tree has been criticised but no other method has yet to emerge which produces more consistent results. The problem is circular in that a good tree is needed to discuss evolution and yet firm ideas are needed about what has occurred in the past in order to construct a truly representative tree. The words of Father Jacobus\textsuperscript{35} eloquently describe the problem faced by evolutionists, “to study history one must know in advance that one is attempting something fundamentally impossible and yet necessary and important. To study history means submitting to chaos and nevertheless retaining faith in order and meaning.” One author\textsuperscript{21} describes the situation as a “tree obscured by vines” indicating the muddling aspect of HT.

Rigorous classification definitions often associated with phylogenies are of human design and of course do not have any effect on the evolutionary processes that actually occurred. Such static ontologies may be replaced in the future by more flexible dynamic ontologies\textsuperscript{20}. Indeed Doolittle\textsuperscript{31} asks, “To what extent is our desire to look at early evolution in terms of cellular lineages preventing us from seeing that it is about genes and their promiscuous spread across taxonomic boundaries, which then have no permanent significance?” The same author predicts that new consensus trees and classifications will eventually emerge that retain cellular lineages but cope more adequately with possible HT events.
References

35. www.ucmp.berkeley.edu/inde.html