Mobile Genetic Elements 2:1, 59–62; January/February 2012; © 2012 Landes Bioscience

Phylogenomic approaches underestimate eukaryotic gene transfer

Jan O. Andersson
Department of Cell and Molecular Biology; Science for Life Laboratory; Uppsala University; Uppsala, Sweden

Phylogenomic approaches have shown that eukaryotes acquire genes via gene transfer. However, there are two fundamental problems for most of these analyses; only transfers from prokaryotes are analyzed and the screening procedures applied assume that gene transfer is rare for eukaryotes. Directed studies of the impact of gene transfer on diverse eukaryotic lineages produce a much more complex picture. Many gene families are affected by multiple transfer events from prokaryotes to eukaryotes, and transfers between eukaryotic lineages are routinely detected. This suggests that the assumptions applied in traditional phylogenomic approaches are too naïve and result in many false negatives. This issue was recently addressed by identifying and analyzing the evolutionary history of 49 patchily distributed proteins shared between Dictyostelium and bacteria. The vast majority of these gene families showed strong indications of gene transfers, both between and within the three domains of life. However, only one of these was previously reported as a gene transfer candidate using a traditional phylogenomic approach. This clearly illustrates that more realistic assumptions are urgently needed in genome-wide studies of eukaryotic gene transfer.

Keywords: protists, phylogenetics, lateral gene transfer, horizontal gene transfer, genome evolution, adaptation, diversification

Submitted: 01/20/12

Revised: 02/06/12

Accepted: 02/08/12

http://dx.doi.org/10.4161/mge.19668

Correspondence to: Jan O. Andersson;
Email: jan.andersson@icm.uu.se

Transfer of genetic material between different organismal lineages is important in prokaryote evolution. Studies of single gene families as well as phylogenomic studies in the last decade have shown that also eukaryotes are affected by this evolutionary mechanism.[1-3] However, the importance of the process is still uncertain; only modest numbers of gene transfer candidates are typically reported from eukaryotic genome projects,[4,5] whereas directed studies suggest gene transfer to be important in the adaptation process of eukaryotes.[6-8] This is an intriguing incongruity. To understand these differences it is useful to consider the assumptions applied in the screening procedures in the phylogenomic approaches used in genome projects (referred to as 'traditional phylogenomic approaches' herein) and how well they match the knowledge we currently have from more directed studies of eukaryote gene transfer. Here I will argue that the match is really poor leading to a high number of false negatives.

Studies of gene transfers in eukaryotes using phylogenomic methods typically identify eukaryotic proteins with high sequence similarity to a prokaryotic protein, but with no or significantly weaker similarity to any eukaryotic protein.[4,5] This is indeed a strong indication of a gene transfer event. The problem is that these traditional phylogenomic approaches only identifies protein families in which a single transfer has occurred between a prokaryote and a eukaryote, which probably is a rarity. Comparative genomics studies of prokaryotes have shown that most protein families are patchily distributed; they are absent from a small or large fraction of the genomes (**Fig. 1**).[9] These genes are distributed via gene transfer and provide diversifying functions and niche adaptation to the recipients. Eukaryote genome evolution, on the other hand, has been viewed as mainly influenced by genome expansion and a few major endosymbiotic events. However, there are data suggesting that gene transfer of patchily distributed proteins is important for the diversification process also for eukaryotes.[1-3]

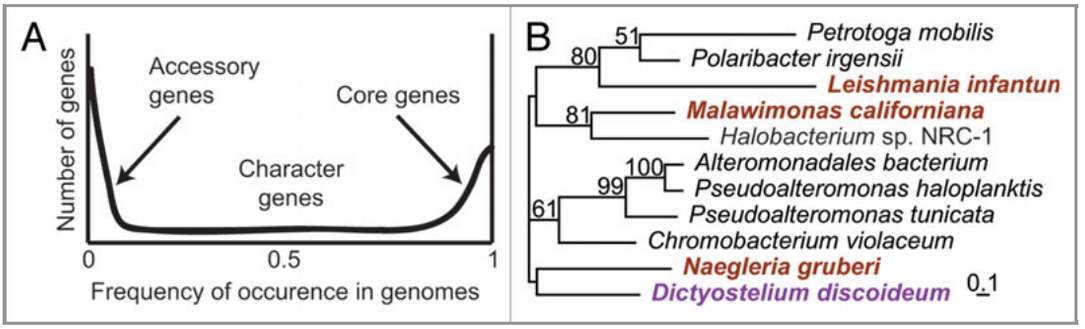

Figure 1. Patchily distributed proteins are distributed via gene transfer. (A) Comparative genomics of prokaryotes have identified three loosely defined groups of gene families based on their frequencies in genomes: extended core, character and accessory genes.[9] Core genes encode shared function between organisms, character genes functions that distinguish major groups and accessory genes functions unique to a few organisms. The evolutionary mode differs between the groups. Core genes are vertically inherited and used for organismal phylogenies, whereas character and accessory genes probably are more influenced by gene transfer; none of the groups should be viewed as representatives of the evolution of the whole genome. The evolution of the accessory and character genes were studied by identifying 49 patchily distributed protein families present in the cellular slime mold *Dictyostelium discoideum* and bacteria.[8] (B) The maximum likelihood phylogeny of a conserved hypothetical protein identified in the study. Eukaryotes are shown in color and prokaryotes in black. Distantly related eukaryotes are found intermixed with prokaryotic sequences, suggestive of multiple transfer events.[8] The figure is adapted from references 8 and 9.

Here I will review some recent findings of adaptation by gene acquisition in eukaryotes obtained by the usage of phylogenomic approaches for the study of gene transfers between specific eukaryotic groups,[6,7] and one attempt to use a novel approach to study patchily distributed proteins.[8] With an increased understanding of the evolutionary dynamics of all classes of gene families (**Fig. 1**) we can apply realistic assumptions to large-scale studies of eukaryotic gene transfer.

Directed Studies Identify Gene Sharing Leading to Adaptation

Many of the most devastating diseases in plants are caused by fungi or oomycetes. These are two distantly related groups of eukaryotes that have similar lifestyles. Both feed by osmotrophy. The cells secret enzymes that decompose organic matter and the metabolites are imported into the cell. The similarity in lifestyle between the groups is an example of convergent evolution. Fungi are more closely related to animals than to oomycetes, whereas diatoms, a group of photosynthetic algae, are a sister group to oomycetes. Absence of phagotrophy has been assumed to be a barrier to gene transfer. Indeed, the oomycete and fungi genomes are not among the genomes for which traditional phylogenomic studies have indicated a significant role of gene transfer in eukaryotes. Nevertheless, targeted evolutionary studies have suggested that gene transfer contributed to the similarities between the groups.[10,11]

Richards and coworkers studied this phenomenon further.[7] They could identify dozens of gene transfers between the groups using a wide range of genomes from both groups together with clustering and phylogenetic methods. Interestingly, all transfers except one were reported to have occurred in the direction from fungi to oomycetes. Many of the transferred genes encode secreted decomposing enzymes and were specifically acquired by plant-tissue colonizing oomycetes. These results show that oomycete most likely are more recent plant pathogens than fungi and that transfer of genetic material from a distantly related eukaryotic group have played an important role in evolution of their pathogenic lifestyle.[7] These fascinating results would not have been obtained with a traditional phylogenomic approach in which genes with strong sequence similarities to other eukaryotes would have been assumed to be present in the common eukaryotic ancestor.

Studies of gene transfer are indeed able to shed light on the diversification process of eukaryotes. Animals and fungi are both members of Opisthokonta. No photosynthetic member has been identified in this group. Choanoflagellates are a group of free-living microbial eukaryotes which are the closest relatives to animals. Sun and coworkers used a directed phylogenomic approach to search for genes of algal origin in the genome of *Monosiga brevicollis*, a phagotrophic unicellular choanoflagellate.[6] They reconstructed phylogenetic trees for all genes in the genome. Using realistic filtering criteria they were able to identify 103 genes with strong support for algal origin, mostly from haptophytes, diatoms and green algae. This could be the result of repeated transfer of genes from food; choanoflagellates feed on bacteria and other eukaryotes. Alternatively, or rather in addition, the genes could have been introduced from a past algal endosymbiont in the lineage leading to Monosiga.[6] Interestingly, a quarter of the identified genes appeared to first have been transferred from bacteria to a eukaryotic alga, and then secondarily to choanoflagellates.[6] However, such a bacterial origin would not have been detected using a typical phylogenomic approach since the strongest sequence similarity would be to algal gene of bacterial origin. Functions in amino acid and carbohydrate metabolism dominated among the gene transfer candidates, indicating that these choanoflagellates have adapted by acquisition of algal genes that expand their metabolic repertoire.

A Novel Approach to Study Patchily Distributed Proteins

The two examples outline above test well-defined hypotheses about gene transfers by using directed phylogenomic methods

in combination with careful filtering and interpretation of the results. They are very powerful to characterize the role of gene transfers between distantly related eukaryotic groups in the adaptation of eukaryotes. However, one disadvantage is that the approach relies on existing knowledge of the biology of the organisms; if only transfers between two organismal groups are addressed important contributions from other groups may be missed in such phylogenomic approaches. In addition, these kinds of studies can only estimate minimal number of transfers between the groups analyzed; the number of false negatives may be large. To circumvent these problems I applied an alternative approach to study these issues. I first identified patchily distributed proteins because these are expected to be enriched with gene transfer events (**Fig. 1**), instead of screening for unexpected sequence similarities. Then I performed phylogenetic analyses for each identified gene family to evaluate whether the patchy distribution was a consequence of gene transfer, or differential loss in the eukaryotic domain.[8]

The soil-dwelling cellular slime mold *D. discoideum* was selected in the case study for two reasons: an active research community have produced a high quality annotation of the genome sequence (http://dictybase.org/), and only 18 potential gene transfers were reported in the original publication.[8] I identified 49 protein families in the Dictyostelium genome which were shared with at least one prokaryotic species, but only a limited number of other eukaryotes and prokaryotes (**Fig. 1**). The evolutionary history of these patchily distributed families were analyzed further.[8] For seven of the families there were no eukaryotic sequences except the Dictyostelium sequences. The remaining 42 families contained sequences from one or more eukaryotic species outside the Dictyostelium genus. The closest relative with a completely sequenced genome, the human parasite *E. histolytica*, was represented in only two families. In contrast, the amoeboflagellate *Naegleria gruberi* had a representative in 25 of the families. Dictyostelium and Naegleria are having somewhat overlapping lifestyles, they are both free-living heterotrophs that can be found in soil and they both undergo cell differentiation under certain conditions. However, they are distantly related eukaryotes classified within two different supergroups: Amoebozoa and Excavata.

There exist at least two alternative plausible explanations for this striking gene-sharing pattern. These genes were present in the common ancestor of the Dictyostelium and Naegleria and distributed in eukaryotes strictly by vertical inheritance.[8] In lineages that have different lifestyles (i.e., parasites) the genes have become obsolete and lost over evolutionary time. Alternatively, the genes have been distributed via gene transfer in more recent evolutionary timescales providing selective advantage to the recipient lineages. Phylogenetic analyses were performed on all protein families to distinguish between these alternatives. The results were striking. The vast majority of the phylogenetic trees showed strong indications of lateral gene transfer between prokaryotes and eukaryotes and within eukaryotes.[8] **Figure 1B** shows an example of an individual gene tree. The exact details of the transfer events could in many cases not be traced, because the density of organismal sampling was too low. Nevertheless, there are no strong indications that any of the proteins have evolved solely via vertical inheritance and gene loss; gene transfer has likely affected all patchily distributed genes families identified in the analysis to some extent.[8]

Traditional Phylogenomic Studies have Drastically Underestimated the Amount of Gene Transfer

Only a single protein among the 49 identified as patchily distributed was among the 18 gene transfer candidates in the original *D. discoideum* genome publication,[4] and very few were among the 184 lateral gene transfer candidates reported from *N. gruberi*.[5] This may be surprising, but is logical if the details of the methods applied are considered. Dictyostelium genes with significant similarity to a bacterial-specific Pfam domain and only present in Dictyostelium among eukaryotes were considered as gene transfer candidates.[4] This conservative approach is unlikely to pick up false positives, but will be very prone to false negatives. Genes acquired via gene transfer in two or more different eukaryotes are excluded, as are any genes without sufficient sampling among prokaryotes to be included in Pfam. Similarly, the *N. gruberi* gene set was screened with similarity searches, and genes with significant similarity only to prokaryotes were considered as gene transfer candidates.[5] Again, gene families with repeated transfers are missed in the screen and eukaryote-to-eukaryote transfers are not even considered. The true number of gene families in these microbial eukaryotes are likely much larger than has been reported.

These discrepancies should not be surprising from a biological viewpoint. Microbes live in steadily changing environments. Ecosystems are inhabited by distantly related organisms which have adapted to its specific condition. The spread of patchily distributed genes are part of this adaptation process, and there is no reason to assume that microbial eukaryotes do not take part of this flux of genetic material (**Fig. 1**).[1,2,6-8] For example, if a gene provide the ability to utilize a carbon compound present in the environment it is likely to spread to different microbes in the environment previously lacking this ability (provided that there are mechanisms in action). The assumption that a gene has a vertical eukaryotic history is violated as soon as two eukaryotic lineages inhabit a similar environment and acquire their copy of a particular gene family independently during the adaptation process. The traditional phylogenomic approaches will fail to identify members of such protein families as gene transfer candidates because they assume that vertical inheritance is the norm for all protein families with gene transfer events as very rare exceptions. However, this is probably only the case for universal core genes, and certainly not for patchily distributed proteins (**Fig. 1**).[8] Traditional phylogenomic approaches probably only have scratched the surface of the gene transfer events and thereby drastically underestimated the impact of the process on eukaryotic genome evolution.

References

1. Andersson JO. Gene transfer and diversification of microbial eukaryotes. Annu Rev Microbiol 2009; 63: 177-93; PMID:19575565; http://dx.doi.org/10.1146/annurev.micro.091208.073203
2. Fitzpatrick DA. Horizontal gene transfer in fungi. FEMS Microbiol Lett 2011; PMID:22112233; http://dx.doi.org/10.1111/j.1574-6968.2011.02465.x
3. Dunning Hotopp JC. Horizontal gene transfer between bacteria and animals. Trends Genet 2011; 27:157-63; PMID:21334091; http://dx.doi.org/10.1016/j.tig.2011.01.005
4. Eichinger L, Pachebat JA, Glöckner G, Rajandream MA, Sucgang R, Berriman M, et al. The genome of the social amoeba *Dictyostelium discoideum*. Nature 2005; 435:43-57; PMID:15875012; http://dx.doi.org/10.1038/nature03481
5. Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, et al. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. Cell 2010; 140:631-42; PMID:20211133; http://dx.doi.org/10.1016/j.cell.2010.01.032
6. Sun G, Yang Z, Ishwar A, Huang J. Algal genes in the closest relatives of animals. Mol Biol Evol 2010; 27:2879-89; PMID:20627874; http://dx.doi.org/10.1093/molbev/msq175
7. Richards TA, Soanes DM, Jones MD, Vasieva O, Leonard G, Paszkiewicz K, et al. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Proc Natl Acad Sci U S A 2011; 108: 15258-63; PMID:21878562; http://dx.doi.org/10.1073/pnas.1105100108
8. Andersson JO. Evolution of patchily distributed proteins shared between eukaryotes and prokaryotes: *Dictyostelium* as a case study. J Mol Microbiol Biotechnol 2011; 20:83-95; PMID:21430389; http://dx.doi.org/10.1159/000324505
9. Lapierre P, Gogarten JP. Estimating the size of the bacterial pan-genome. Trends Genet 2009; 25:107-10; PMID:19168257; http://dx.doi.org/10.1016/j.tig.2008.12.004
10. Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. Curr Biol 2006; 16:1857-64; PMID:16979565; http://dx.doi.org/10.1016/j.cub.2006.07.052
11. Belbahri L, Calmin G, Mauch F, Andersson JO. Evolution of the cutinase gene family: evidence for lateral gene transfer of a candidate *Phytophthora* virulence factor. Gene 2008; 408:1-8; PMID:18024004; http://dx.doi.org/10.1016/j.gene.2007.10.019