Annu. Rev. Genet. 1994. 28:237–61

HORIZONTAL GENE TRANSFER: Evidence and Possible Consequences

M. Syvanen
Department of Medical Microbiology and Immunology, School of Medicine, University of California, Davis, California 95616

KEY WORDS: phylogeny, mosaics, angiosperm, molecular evolution, transgenic organisms

CONTENTS

> As careful observation is far harder work than generalization, and still harder than speculation, do you think it very possible that it may be overvalued.
> From C. Darwin to Asa Gray, 1857.

INTRODUCTION

The possibility that genetic information can move between remotely related species is an idea that has been strenuously resisted within traditional schools

0066-4197/94/1215-0237$05.00

of biology. This is especially true of evolutionists who have developed neo-Darwinian theory. This resistance is natural, since superficially the concept seems at odds with the explanatory power of the phylogenetic tree in taxonomy and the important role of reproductive isolation as a mechanism of speciation. At the same time, progress in the field of molecular genetics over the past 25 years has led to many different observations consistent with horizontal gene transfers. It is not just that there are dozens of examples of likely horizontal transfers, but that the mechanisms which could accommodate horizontal transfer of DNA are observed everywhere. Multiple mechanisms for the physical transfer of DNA from one species to another are known. Recombination mechanisms that can absorb this DNA are ubiquitous. Examples of cells and organisms that will express and incorporate products from foreign genes are too numerous to list. Indeed, the concept of genetic coadaptation, which was a seriously accepted idea 10 years ago, is now rarely cited. Coadaptation posits that genes within a genome become so specifically coadapted with one another that they will fail to function with alternative alleles. The many cases of transgenic animals and cell lines created in vitro renders this idea obsolete. The only remaining question is whether the horizontal transfer of genes occurs at a rate that significantly influences evolution.

In 1982, when I first theorized that horizontal gene transfers are likely a major evolutionary force (91), my reasons were twofold: First, it had recently been established that the mechanisms for gēne transfer were present at a molecular level, and my intuition told me that if a mechanism so potentially useful existed, nature would find a way to use it. Second, a general evolutionary theory incorporating the idea of the migration of genetic information across taxonomic boundaries seemed to provide a simple and satisfying answer to a question seldom asked: Why is the molecular biology for all living organisms so unified? In the face of speciation, biology has retained a unity so profound that transgenic animals can be created in the laboratory—a phenomenon that few would have predicted in 1982. In fact, most evolutionists still believe that a theory of evolution incorporating horizontal transfer conflicts with the useful notions of phylogenetic trees and reproductive isolation. In the following discussion on the nature of the evidence supporting horizontal transfers, I try to resolve this apparent contradiction.

When I first discussed the theoretical implications of horizontal transfer, there were less than half a dozen suspected examples involving eukaryotes in the literature (90, 91); only one of these has held up under closer scrutiny. Yet reports of suspected cases continue to grow, and many appear to be serious candidates. Because so many examples have been offered and then retracted, and because there remains considerable skepticism as to the occurrence of horizontal transfer, I examine some of the criteria for identifying gene migrations when they have occurred, and I consider the rigor of the various criteria

for making this judgment. I then explore some old problems that concern flowering plants. In this case I believe transfers may be too frequent to apply criteria that depend on knowledge of species phylogeny. I try to document the crisis in angiosperm systematics and paleontology between neoDarwinian theory and its database—a crisis that could be resolved by a theory of evolution incorporating horizontal gene transfer.

The topic of horizontal transfer overlaps with the endosymbiotic theory of organelles. Endosymbiosis basically involves the fusion of the entire genomes of two organisms. I am considering these to be one part of the larger phenomenon of cross-species gene transfer, which involves, in addition to endosymbiotic fusion, the insertion of smaller genetic regions, including single genes or even parts of genes. The mechanisms of transfer will likely involve viruses (30), direct transformation, conjugation, or other as yet unexplored means. This review does not focus on the possible mechanisms for DNA penetration through species boundaries.

With the growth of the DNA database, especially of large homologous sets, the subject of horizontal gene transfer has grown quickly and has been recently reviewed (38, 82, 86).

TRANSFERS INVOLVING EUKARYOTES

Identifying horizontal transfers that have occurred in the recent past, such as the spread of antibiotic resistance plasmids, is not problematic. The difficult cases are those that have occurred so far in the past that subsequent divergence of the gene sequence may have obscured the original relationships. It is on these cases that I concentrate.

Phylogenetic Congruency Test

The most rigorous criteria available for establishing the occurrence of ancient horizontal gene transfer involve comparing the phylogenetic tree constructed with regard to a specific protein or gene from a number of distantly related organisms with that of the known phylogeny for those species. If one member belongs to a radically unexpected assemblage, i.e. if an incongruency is seen between the "gene tree" and the "species tree", then a case can be made for a horizontal transfer. This test, which has been used with varying degrees of rigor since the earliest claims of horizontal transfer, has been most explicitly defined by Smith et al (82).

Ideally, application of the phylogenetic congruency test requires that a number of conditions be fulfilled. First, the genes being considered must have phylogenetic information. Second, the homology being compared must represent orthologous, but not paralogous, genes (see below). And third, the incongruent example should be a gene whose substitution rate is not radically

Table 1 Some horizontal transfers involving eukaryotes

Protein	Case	Reference
1. Glucose phosphate isomerase	*gpi* from *E. coli* was inherited from a relative of the dicot *Clarkia ungulata*	82
2. Fe superoxide dismutase	The Fe SOD from the protist *Entamoeba histolytica* has prokaryotic origins	82
3. Aldolase	Class II aldolase from yeast shows affinity to that from *E. coli*	82
4. Cytochrome c	The protein from *Arabidopsis thaliana* is of fungal affinity	37
5. Xylanase	*Rumoncoccus*-like gene is found in a rumen fungus	23
6. Thioredoxin	Thioredoxin-m in plants has a bacterial origin	31
7. Glyceraldehyde-3-phosphate dehydrogenase	*gapdhA* from *E. coli* and *Anabaena* have an affinity with eukaryotic *gapC*	16, 54
8. Elongation factor Tu	*tufA* in *Arabidopsis* is from its endosymbiont	4
9. Ribosomal proteins L21 and L22	L21 and L22 in some plants is from its endosymbiont	22, 55

These examples all involve nuclear genes in eukaryotes. Examples 1–7 were indentified by the phylogenetic congruency test and 8 and 9 were inferred from finding Tu, L21, and L22 in both nuclear and chloroplast genomes among contemporary plants.

different from the other genes being compared. Before these three conditions can be assured, the number of genes being compared needs to be reasonably large, i.e. more that 5 and possibly more than 10. Table 1 summarizes examples that involve eukaryotes, including those reviewed by Smith et al. The phylogenetic congruency test supports the hypothesis of lateral transfer in seven of the nine cases. All of these examples involve a complete nuclear gene. Examples 6–10 are probably bacteria-to-eukaryotic transfers that involved mitochondrial or chloroplast intermediates, i.e. they were part of the original endosymbiosis that gave rise to those organelles; there are no clues as to the mechanism of transfer for the other six.

In some of the examples in Table 1, the direction of transfer is given. Logically, directionality is difficult to infer from the phylogenetic congruency test, especially for ancient transfers involving lineages that gave rise to many extant taxa. An example of this difficulty is illustrated by glyceraldehyde

3-phosphate dehydrogenase (line 7, Table 1). Doolittle et al (16) concluded that *gapdhA* in *Escherichia coli* came from a eukaryotic source, since it was the sole prokaryote in a eukaryotic clade. But with the finding of an orthologous *gapdhA* in *Anabaena,* it seems more likely that an ancestral bacterium donated *gapC* to the eukaryotes (54).

Pitfalls Involving Paralogies

Most of the premature reports of possible horizontal transfers resulted from comparing paralogous genes and treating them as if they were orthologous. Two genes are defined as paralogous if they diverged after a gene duplication, whereas the two genes are defined as orthologous if they diverged after a speciation event. Let us construct the following scenario to illustrate this problem. Consider a gene in an ancestral strain that underwent a duplication to give rise to gene A and gene B, which subsequently diverged from one another. Further consider the evolution of that ancestral strain into two major branches. In the extant species, all descendants of gene A will belong to one orthologous set, and descendants of gene B will belong to another. Members within the gene A group are said to be paralogous to members of the gene B group. If both sets can be identified in the extant species, this does not pose a problem. However, consider what happens if, during the evolution of these two major branches, the duplicated pair is reduced back to a single copy, and A survives in some extant species while B survives in others. This could result in both gene A and gene B being located in the two major clades and their paralogous nature not being recognized. If a single tree is constructed, then the A-B gene tree would be incongruent with the species tree and hence a horizontal transfer incorrectly inferred. This scenario can also arise from what is essentially a sampling error, when too few genes and their incompletely characterized products are compared.

Comparisons of paralogous genes were responsible for the early report that the Cu-Zn superoxide dismutase in the bacterium *Photobacterium leiognathi* came from its eukaryotic host, and that the plant leghemoglobin came from vertebrates. As more sequences were analyzed and orthologous genes identified, it was found that the respective gene trees became reasonably congruent with their species trees (6, 82). Another controversial example has been the possible eukaryotic origin of glutamine synthetase II found in plant symbiotic bacteria (78). The most recent analysis of this example (82) favors a very ancient horizontal transfer, though this is likely not the last word.

Also, the recent suggestion that cytochrome p450 in *Fusarium oxysporum* transferred from a *Streptomyces* ancestor (40) suffers from the problem that orthologous families of cytochrome P450s are extremely difficult to identify. The extent of this problem is illustrated by the complete cytochrome P450 gene tree constructed by Degtyarenko & Archakov (14) who, though they

conclude horizontal transfers and "convergent" replacements likely occurred, were unable to relate this tree to any underlying phylogeny. It would appear that the multiple horizontal transfer of cytochrome P450 is not an unreasonable hypothesis, we must conclude for the present that its phylogenetic "tracks" are indecipherable.

Pitfalls Involving Unequal Rates

In small data sets, unequal rates of substitution may not be apparent. As noted by Felsenstein (20a), comparing genes experiencing grossly unequal substitution rates can result in aberrant affinities during phylogenetic reconstruction. This problem was encountered in striated muscle calmodulin in the chicken—a second calmodulin in the chicken that, at the time of its discovery, had no counterpart. Gruskin et al (28) showed that this gene, designated calmodulin-like (*cl*), was highly diverged from the other chicken calmodulin gene (*cam*), as well as from all known vertebrate and mammalian genes. From this fact, and from third base similarities between *cl* and the electric eel calmodulin, they suggested that *cl* entered the chicken by a horizontal transfer, possibly by a viral mediated retrotransposition (because it is an intronless gene).

The case for a migratory chicken calmodulin was inferred from the minimal replacement tree based on the half dozen available sequences. Today, there are over 60 different sequences (66). With this database, we can reanalyze the likelihood that the chicken cl gene resulted from a horizontal transfer. Figure 1 gives the phylogeny of a set of 18 calmodulins from various eukaryotic sources. The length of the branches in the tree in Figure 1 gives a rough measure of the number of replacements since common ancestors. This tree

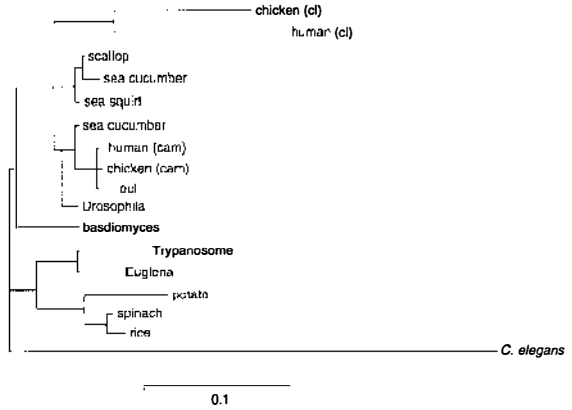


Figure 1 Phylogenetic tree based on the protein sequence of calmodulin from a variety of eukaryotes. This tree is determined from a hybrid of procedures—the branching order was determined from the minimal replacement method, and the length of the branches corresponds roughly to an average distance between taxa on opposite sides of each node.

reveals a number of points. One is that the rate of change of the chicken *cl* gene is much higher than that of the other vertebrate calmodulins. Indeed, the branch labeled human (cam) includes calmodulins from rat, mouse, bovine, xenopus, and salmon, indicating that the gene is completely functionally constrained among all these vertebrates. In addition, a human analog to *cl* is now known that clusters with the chicken cl. Hence, there are apparently two different calmodulins: one is the highly functionally constrained member found in most eukaryotes; the other is a less constrained member found in human and chicken. The chicken *cl* gene may not be orthologous to the other members to which it was originally compared, and almost certainly is evolving at a much higher rate than the others. Thus, the chicken *cl* gene fails on possibly two of the three conditions that we set out and, therefore, can no longer be considered to have experienced a horizontal transfer.

Convergent Replacements

Any discussion of horizontal gene transfer must deal with the possibility of convergent evolution. Some scientists still do not accept horizontal gene transfer as the explanation for an unexpected association that is revealed by the phylogenetic congruency test; rather they attribute this association to convergent replacements. For example, Kemmerer et al (37), in their paper showing the affinity of cytochrome c in Arabidopsis with those of fungi, do not offer a mechanistic explanation. In a recent review of this problem (15), Doolittle concluded that although functional and mechanistic convergence is common and enzyme structural convergence has probably occurred, no convincing case for genuine sequence convergence has yet been made. In the most widely cited example of sequence convergence, that of the langur lysozyme converging toward true ruminants (89), Doolittle showed that the lysozyme tree is congruent with the species tree. That is, the convergent amino acid replacements that occurred in the lineage leading to ruminants and the langur were too few in the background of neutral changes to hide the affinity of the langur's lysozyme with that of the other primates. In a similar example, we encountered convergent changes in a bacterial neomycin phosphotransferase that may be the result of natural selection, but the number of these changes was too low to distort gene phylogeny (48).

The lack of evidence that convergent evolution at the sequence level is responsible for discrepancies between species trees and gene trees has produced a subtle shift in evaluating evidence for horizontal transfers. Fifteen years ago, because the notion of horizontal gene transfer was so completely unexpected and apparently contrary to the theory of the biological species, evidence in its favor would possibly have been attributed to convergent evolution. Today, with numerous mechanisms now known for moving genes through species barriers, and with the better documented cases, major in-

congruencies between species and gene trees are increasingly being explained by horizontal transfer.

Tests of Significance

To say we have found an "unexpected" assemblage from the phylogenetic congruence test implies that statistical expectations of outcome should be computed, thereby allowing us to calculate a significance for our "unexpected" result. Unfortunately, rigorous statistical tests are not generally available for testing the significance of phylogenetic trees. The computational problems are immense (32). Indeed, for data sets involving large numbers of taxa, it can be exceedingly difficult to even find the shortest tree, let alone test the significance of that tree against another. This problem has attracted much interest recently, and approaches for calculating the significance of competing four and five taxa trees have been suggested. These include Felsenstein's (21) and Kishino & Hasegawa's (39) maximum likelihood methods, and Lake's (46) method of maximum parsimony. The difficulty with the maximum likelihood procedures is that before a significance test can be computed, a model of the evolutionary process must be assumed. This model usually assumes that both replacements along lineages and branching events follow a Markov process. However, it is one thing to ask if a given tree is consistent with a specific model, and quite another to ask if it can discriminate between two competing models. These new procedures are being used to test the significance of phylogenetic incongruencies, but as yet these approaches have not been applied to testing the significance of horizontal transfers.

Constructing phylogenetic trees from sequence data that contain "homoplasy" can make the congruency test even more difficult. Homoplasy arises when distantly related species share unique traits. The difficulty lies in distinguishing whether these shared traits reflect inheritance from a common ancestor, or whether they were derived independently. Traditionally, homoplasy is considered to result from processes such as convergent and parallel evolution and reversion to ancestral states; obviously, characters within a data set involved in horizontal transfers would also contribute to homoplasy. The implications of this are described in greater detail in the discussion of the angiosperms.

MOBILE GENETIC ELEMENTS

If the phylogenetic congruency test is applied in the absence of other evidence, it can be too restrictive. It may preclude from consideration genes that are frequently involved in horizontal transfers. This is likely the case for many transposable elements—horizontal transfer is so frequent for some elements that the phylogeny of the species carrying these elements is lost. Kidwell (38)

recently reviewed this subject in detail, so I briefly touch on just a few examples. In the examples of horizontal transfers of eukaryotic genes, the first and still one of the most convincing cases is the P-factor in *Drosophila melanogaster*. This case is convincing because the transfer event occurred in recent years and was therefore observed as it happened. Observing the event in natural populations in real time makes for a very convincing case, much as observing the spread of antibiotic genes on plasmids among pathogenic bacteria has firmly established the role of plasmids in the horizontal spread of traits in bacteria. The P-factor is convincing for a second reason. Houck et al (34) showed that the mite *Proctolaepus* is a possible donor of the P-factor to the *Drosophilids*. In addition, the phylogenetic congruency tests convincingly support horizontal transfer (13). Since the discovery of P-factors, numerous examples involving other transposable elements in *Drosophila*, as well as elements in plants, yeasts, and other metazoans, have been recorded. In some of these examples, many of the possible transfer events occurred in the distant past. Subsequent sequence divergence has so obscured the affinities that reconstruction of their evolutionary histories is either impossible or highly conjectural. The phylogenetic history of many mobile elements looks much like the phylogeny of viruses (as some are mixed in with viruses) in that their history is relatively independent of the phylogeny of their respective hosts (104).

The two classes of sequences of highly repetitive DNA found in vertebrates, the SINEs and LINEs, have attracted considerable attention because they are abundant in the human genome. Dispersion of both is likely due to retrotransposition events, with the LINEs appearing like common, genomic transposable elements and the SINEs having some retroviral affinity. On the basis of the phylogenetic congruency test, both elements have probably transferred horizontally, at least into some of the taxa that now carry them (65, 68).

CHROMOSOMAL GENES IN BACTERIA

The subject of bacterial plasmids and their associated accessory DNAs has been extensively reviewed recently (1, 49, 60) and thus is not covered here. It has been known since Griffith discovered transformation in 1928 that the mechanisms for the horizontal transfer of information are in place. That this mechanism acts on chromosomal genes has been known since the 1950s. The only remaining question has been: Does this happen in nature? This has been a difficult question to answer.

Mosaic Nature of the E. coli *Chromosome*

Until very recently it was even questioned whether genetic information transferred between different strains of the same *E. coli* species. It is useful to

review this history because the logical structure of its discussion has mirrored the larger discussion of cross-species gene flow. In particular, two issues have emerged. First, the argument has been made that horizontal gene transfer cannot be important in *E. coli*, because if it had occurred, strain differences would have been obliterated. Second, the process of answering this question is leading to new criteria for recognizing horizontal transfers between closely related groups.

Selander's group focused on natural populations of *E. coli* and concluded that its population structure was "clonal." From this, he concluded that gene flow and recombination between natural *E. coli* strains must not be important (67, 101). This widely cited conclusion was based on the finding that native populations of *E. coli* could be divided, based on a genetic distance derived from enzyme polymorphisms, into at least three different groups in which members of a group were more closely related to each other than to members of the other groups, i.e. *E. coli* populations can be divided into demes.

The notion that horizontal transfer among *E. coli* was absent because of its clonal population structure was dropped after a number of comparisons were made between extensive DNA sequences derived from different strains. This analysis was initiated by Sawyer (76), but Milkman & Bridges (63, 64) have presented the most rigorous analysis. They determined the complete nucleotide sequence of a 4400-bp region of the *trp* operon from 36 strains of *E. coli* selected from the same set of strains that Selander's group used to determine the "clonal" population structure. They confirmed the earlier finding that the 36 strains could be divided into essentially the same groups identified by enzyme polymorphisms. In addition, however, they found when comparing strains within a group that one of the members may have a short section that differs from that of the other members. That is, in the idiosyncratic individual regions of similarity are occasionally interspersed with regions of dissimilarity. Moreover, the region of dissimilarity could often be found in one of the other groups, as if this region of DNA had transferred from one to the other. From the average size of the dissimilar regions, it has been estimated that the average recombination event usually results in the transfer of only a few hundred to 1000 base pairs. Pairs of chromosomes with this pattern of similarity interspersed with regions of dissimilarity are said to have a mosaic pattern. A conclusion of a mosaic pattern is based on models whose significance can be rigorously tested (58, 64, 76). In summary, this work clearly demonstrates the presence of subpopulations within *E. coli* that are genetically distinctive, but that occasionally experience a gene transfer from one group into another, i.e. small breaches of a "clone's" reproductive isolation do not destroy its identity.

In general, a comparison of any two homologous DNA regions that display an abrupt change in similarity raises the possibility of a mosaic structure, and hence a horizontal transfer. The case for *E. coli* is solid because the numbers

are sufficiently high that the statistical significance of the discontinuities is high. Furthermore, the donor strains were identified. I have used the criterion of discontinuity in the degree of similarity to identify possible horizontal transfers involving small regions of mammalian beta globin genes (90). In this case, the discontinuity was statistically significant, but it remains to be seen whether it represents horizontal transfer.

Salmonella and the Criterion of GC Content

Considerable progress has been made in characterizing different pathogenic strains of *Salmonella,* and this process has led to numerous claims of horizontal transfers. There are many good examples of horizontal movement of virulence factors. Beltran et al (5) have shown that the common surface antigens used to classify *Salmonella* serotypes are discontinuously distributed across distantly related strains, as determined by enzyme polymorphisms, which indicates that these serotype genes move horizontally. The virulence capsular antigen encoded by the *viaB* gene found in isolates of *Salmonella dublin* is most likely a horizontal transfer. Although the source organism is not known, it is likely another *Salmonella* serovar (77, 83). Hoyer et al (35) found that the sialidase gene, *nanH*, moves between different strains of *Salmonella* and may even have originated within an assemblage that includes the *Clostridia.*

Groisman et al (27) have viewed the chromosome in *S. typhimurium* as a widespread mosaic of remotely related parts. This conclusion for *Salmonella* is based, in part, on the comparison of genes among enterics. Even though *E. coli* and *S. typhimurium* share genomes of comparable size with the same gene order and with 90% of their genes being closely related, about 10% of the genome in *S. typhimurium* encodes functions not found in *E. coli*. In addition, the GC contents in these unique genes are frequently significantly lower than the 50% average for the entire genome, a finding that has led to the suggestion of remote origins for these regions. Examples of low GC content include phoN (26), the *rfb* genes for o-antigen synthesis (75, 98), and an unknown transcriptional regulator (27). These considerations support the notion of a mosaic chromosome for *Salmonella.* On the other hand, the case for remote species donors for these genes is problematical in that the only evidence for remote origins is deviation of the GC content.

Because various different factors probably influence GC content, I examined this point in greater detail with respect to the hypothesis of remote origins. I collected a distribution of the GC content found in 757 different DNA fragments from *E. coli* (Figure 2A) and in 131 fragments from *S. typhimurium* (Figure 2B). For *E. coli*, the distribution can be divided into two groups. Most fragments cluster about a GC content of 0.509. In fact, this main distribution is close to Gaussian, with a standard deviation determined by a stochastic process, as shown. The second class of DNA fragments skew significantly to

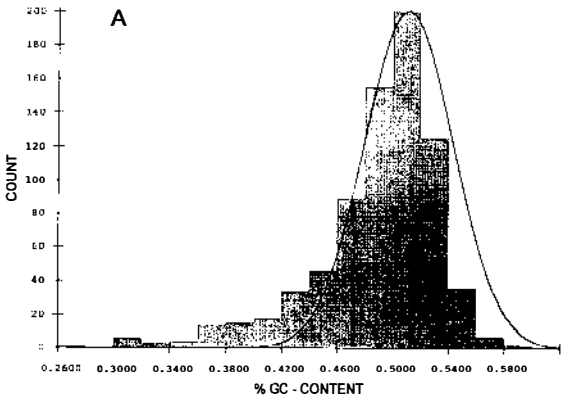
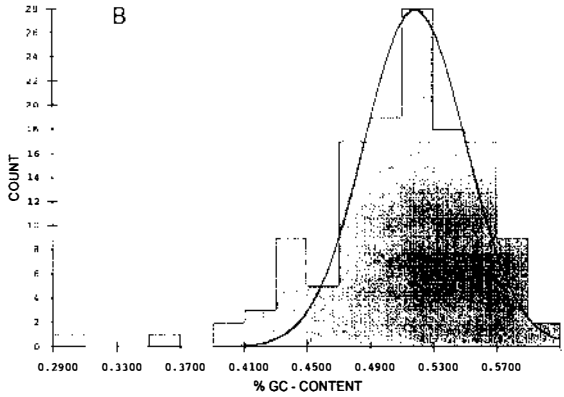


Figure 2 A. Distribution of GC content in DNA fragments from the chromosome in *E. coli*. The sample was selected from the gene bank by listing all genes with Ba:Eco in their entry name and then deleting all entries of plasmids, duplicated entries, and sequences of less than 100 bp. The theoretical curve is the normal, which is computed for a mean of 0.499, and $\sigma = 0.032$.

lower GC content. By the GC-content criterion, these fragments become candidates for horizontal transfers. From the data in Figure 2a, this argument has one problem: the skewing is only to low GC content. The variation to high GC content is consistent with the variance predicted for random fluctuation in GC about a mean of 0.509. Why do we not see genes of high GC-content that came from remote sources?

The case for *S. typhimurium* is not too different. The main clustering near the median of 0.516 is only approximately normal (possibly due to the smaller sample size) and in addition, there may be some skewing to high GC content but, as with *E. coli*, most of the skewing is to the lower GC content. Because the skewing is primarily to low GC content, support for remote species origins for these genes would seem unlikely. Functional selection for low GC content would appear more likely. Obviously, such DNA would have a lowered melting temperature, and it would be easy to imagine replication or recombination scenarios that would select for these genomic regions. For example, a simpler explanation for these low GC regions is that they are accessory DNAs, regions that transfer horizontally among different strains of *Salmonella* and *E. coli*. This model for the *Salmonella* chromosome is analogous to the cassette model for the lambdoid bacteriophages (11), a model that involves recurrent transfer and recombination events of different cassettes due to resulting episodic selection for these accessory DNAs. The lower GC content could thereby be selected at the step of recombination either by easing the melting of DNA, or possibly by reducing the barrier to recombination between heterologous DNAs by reducing the probability of mismatched bases in heteroduplex overlap



B. The distribution of GC content in DNA fragments from the chromosome in *S. typhimurium*. The sequences are from gene bank listings with Ba:Sty in their entry name and edited as were the *E. coli* sequences. The theoretical curve is the normal with a mean of 0.516 and a $\sigma = 0.032$.

regions (74). (The reduced probability of a mismatch at the extremes of GC content can be deduced from a simple statistical argument. Let us compare two randomly generated sequences that are of the same GC content (= g) and compute the probability (P) that at any given site they misalign. Then it can be shown that $P = \{0.5\} + g - g^2$, which reaches a maximum value of 0.75 at $g = 0.5$ and declines to 0.5 as g approaches 0 or 1.)

To conclude, before a remote origin for these regions from the Salmonella can be accepted, a homologous gene from a member of the donor clade must be identified.

Codon Bias

Another criterion for identifying horizontal transfers is to compare codon usage of a gene with the codon bias of the host organism. This is a difficult criterion to apply because codon usage for minor proteins deviates from the bias. Medigue et al (61) examined 740 genes from *E. coli* with respect to codon usage. They found three classes: (*a*) the heavily expressed proteins that define bias in *E. coli*; (*b*) the proteins, expressed in trace amounts, that use the rare codons; and (*c*) a residual group. This third group contained the genes that would have been predicted to be migratory, such as the insertion sequences and other identified mobile elements, and the genes determining surface antigens (as with *Salmonella*, perhaps *E. coli* changes its coat from time to time because of episodic immune selection). But curiously, group (*c*) also contained genes specifying faithful DNA replication, the antimutator genes. This latter fact correlates with the observation that a number of antimutator genes are

plasmid-borne. Perhaps the evolution of bacterial lineages find episodic need for antimutators or, possibly, periodic need for high mutator activities.

Though analysis of codon usage is an interesting exercise in sequence analysis, possible explanations for any deviation from the species bias are sufficiently numerous to preclude its use as a criterion for inferring a horizontal transfer. At best, codon bias can be used to support more compelling evidence.

Other Bacteria

As recently as 1988, it was argued that horizontal transfer contributed little or nothing to the genomes of bacteria. Today, few bacteriologists are eager to push this point—too many examples have been reported during the past five years. Some cases mentioned here are supported by small data sets, and may eventually not hold up as additional sequences are determined. Also, some of these horizontal transfers may involve shuffling cassettes between a closely related group of strains, much as with the examples described above in *Salmonella*. These examples include the *celY* gene in *Erwinia chrysanthemi* (29), fimbrae genes of *Bacteroides nodosus* (33), the CO dehydrogenase from carboxydotrophic bacteria (44), capsule genes from *Haemophilus influenzae* (45), the M12 serotype of streptococci (81, 100), and an amidase from *Brevibacterium* (84). Genes for the photosynthetic reaction center from *Rhodocyclus gelatinosus* (an α purple bacteria) define a clade with its homologous genes from β purple bacteria; this case is interesting historically because these photosynthetic bacteria were once classified, before the days of rRNA trees, into one group on the basis of similar photosynthetic biochemistry. The nitrogenase Fe protein from *Archaebacteria* (85) and a superoxide dismutase in *Halobacterium* (57) involve archaebacterium and may represent horizontal transfers across great phylogenetic distances.

Mosaic patterns indicative of horizontal transfer have also been observed in antibiotic resistance genes in the genomes of *Neisseria* and *Streptococcus*. Naturally occurring resistance in some of their species is encoded by altered penicillin-binding proteins (PBPs). The pathogenic strains *S. pneumoniae* (17), *N. meningitidis* (9, 50), and *N. gonorrhoeae* (87) were originally sensitive, but have recently become resistant. Comparison of the PBP genes from the resistant and sensitive strains reveals a mosaic pattern—regions of similarity interspersed with regions of dissimilarity. The genes from resistant pathogenic strains consist of a hybrid sequence resembling parts from the sensitive strain interspersed with blocks from the naturally occurring resistant strain (59). Spratt et al (87) have shown that, in *Neisseria,* blocks of a few hundred base pairs were transferred. Radstrom et al (73) have documented a similar mosaic pattern in the dihydropteroate synthase gene found in sulfonamide-resistant *N. meningitidis*.

ORGANELLAR GENOMES

The concept of horizontal transfer involving eukaryotic plastids has a long history. In 1905, Mereschkowsky (62) argued that chloroplasts were bacterial endosymbiotes, and in 1928, Walling (97) made a similar case for the mitochondria. Some 50 years later Margulis (51) reintroduced and developed this idea. Today it is the most widely accepted form of horizontal movement of genetic information across high phylogenetic barriers. The demonstration by Woese & Fox (102) that rRNA from the mitochondria and chloroplasts are more closely related to bacteria than to their eukaryotic nuclear counterparts has provided the most convincing evidence for the endosymbiotic theory. Indeed, their paper represents the first application of the phylogenetic congruency test to establish a horizontal transfer.

But beyond the original endosymbiotic event creating the mitochondria and plastids, these genomes have likely been involved in additional horizontal transfers. For example, the plastid phylogeny, which is based on the rubisco *rbcL* genes, places the nonchlorophyll-B algae with the purple bacteria, whereas their 16S RNA sequences place them with the cyanobacteria and green plant chloroplasts (2, 3, 80, 95). Such conflicting phylogenies have led to the hypothesis that the plastids arose by more than one endosymbiotic event. However, to accept this theory, one must posit many parallel mutations (80); a much simpler explanation is to accept the monophyly of the plastids, but to add subsequent horizontal transfers (3). The finding of multiple genes in paramecium mitochondria that have previously only been seen in plant chloroplast raises the possibility of a major migration across species boundaries (72). This example must await a more complete application of the phylogenetic congruence test before it can be accepted as a horizontal transfer, or at least a horizontal transfer that occurred after the early symbiosis events. At least this finding implies a transfer from chloroplast to mitochondrion (20).

There are numerous accounts of organellar DNA that have moved into the nucleus. These examples are relevant to the current review because they define part of the pathway for endosymbiotic-mediated horizontal transfer. These transfer events are both ancient, as evidenced by thioredoxin-F and GAPDH in Table 1, as well as more recent (reviewed in 10, 19, 25, 69).

THE ANGIOSPERM PARADOX

Classification Problem

Darwin's theory of evolution with modification from common ancestors was rapidly accepted by zoologists in 1859, in large part because metazoan assemblages were so easy to organize into trees, with ancestral branch points repre-

sented by animals that, though primitive, could be still classified with their descendants. Botanists, however, have long had difficulty in classifying the flowering plants (the angiosperms) into taxonomic groups above the family level. The traditional problem has been that morphological characters upon which trees could be constructed give conflicting trees, depending upon how the characters are weighted. Thus, botanists have never been able to reach agreement upon the shape of a species tree (88). Even some of the simplest questions remain unresolved. For example, it is not known whether the angiosperms evolved from a common ancestor that could be classified as an angiosperm (i.e. monophyletic), or whether they arose from more than one nonangiosperm lineage (polyphyletic). Initially, those who thought about the evolution and classification of the angiosperms had trouble accepting Darwinian evolution, at least as it is summarized by the phylogenetic tree. This problem of angiosperm classification was apparent when the analysis was based on morphology. With the introduction of sequence data, the problem has only worsened.

The problem in plant taxonomy lies in the widespread existence of parallel traits; thus, any tree that can be constructed depends on a judgment as to which shared traits are ancestrally derived and which evolved "independently" in different lineages. Indeed, both Vavilov, in 1922, and Went, in 1971, focused on parallel variation as major factors in angiosperm evolution. An early goal of molecular studies of angiosperm evolution was to use sequence information contained in a number of genes in order to determine acceptable species trees. This attempt has so far failed for the simple reason that different genes have yielded different assemblages. The first extensive data were based on the protein sequence determined for cytochrome c from 25 angiosperms by Boulter (7). Since then, plastocyanin (52), the large rubisco subunit, rbcL (12, 53), stable RNA genes (7, 18, 53, 94), and glyceraldehyde-3-phosphate dehydrogenase (56) have been used to determine trees. The assemblages determined from one gene (or group of genes and characters) conflict with those determined by another. In addition, it has been very difficult to find a satisfying congruence between any of the gene trees and those relationships agreed upon by systematic botanists. I would suggest that horizontal gene transfer has been sufficiently common during the descent of the angiosperms that different genes in the same species have likely followed different phylogenetic histories. If so, then the phylogenetic congruency test (at least, as outlined above) cannot be used, since finding an acceptable species tree may well be impossible.

The problem in angiosperm classification goes even deeper than conflicting gene trees. Internally robust trees based on single genes are also elusive: the sequence for a single gene is difficult to organize into a tree-like pattern. This problem of parallel variation within a single gene has been apparent since the earliest introduction of molecular data for plants (70). Hartman, Stevens, and

I (93) showed that the cytochrome c tree based on angiosperm sequences was internally homoplastic as compared to that from vertebrates. In this study, we showed that in many respects the plant cytochrome c appears to have evolved in a well-behaved fashion, not differently from, for example, the vertebrate cytochrome c. The plant sequences (with the exception of *Arabidopsis*) diagnose their own kingdom, and their rate of change, that is their molecular clock, is relatively uniform. In fact, the amount of divergence among the sequences from angiosperm cytochrome c is comparable to that in vertebrates. While the gene tree for the vertebrate cytochrome c is reasonably congruent with accepted phylogeny, the plants, on the other hand, yielded four groups of minimal trees that were both highly incongruent with each other and with any acceptable phylogeny. That is, more homoplasy appears in plant cytochrome-c evolution than in vertebrate cytochrome-c evolution. If this internal cytochrome-c homoplasy is due to horizontal transfer events, the segments of transfer must be considerably smaller than a single gene. This means that not only have DNA fragments moved horizontally, but they must have experienced a gene conversion event, thereby leaving only a block of new information inside a preexisting gene.

Temporal Problem

Martin et al (56) drew attention to a major discrepancy that has emerged between molecular evolution and paleobotany. The molecular distance between glyceraldehyde-3-phosphate dehydrogenase (*gapdA*) sequences suggests that various angiosperm families diverged as early as the Permian period, possibly even the Carboniferous, close to 300 million years ago (MYA), while the fossil record supports an angiosperm radiation in the late Cretaceous period—only about 110 MYA. An early divergence of multiple angiosperm families is also supported by the amount of divergence seen for other sequences. Wolfe et al (103) have shown that this holds for the chloroplast rRNA and rubisco sequences. The same result can be deduced from an analysis of the cytochrome-c sequences (see the molecular distance tree given in Figure 3). Nine different angiosperm lineages appear to have diverged at least in the Triassic period, if not the Permian period. The time axis is calibrated using the usual assumptions of a relatively constant molecular clock where the tree is rooted to the ginkgo and where the time of separation of the lineage giving rise to the ginkgo from that giving rise to the angiosperms is in the upper carboniferous (about 260 MYA, though a time of 300 MYA–320 MYA has been used by some authors). The problem is that the earliest fossils resembling members of those nine lineages are no older than 110 MYA. As described above, it is not easy to attribute this discrepancy to aberrant sequence evolution of cytochrome c. Thus, the cytochrome c sequences, in addition to those from

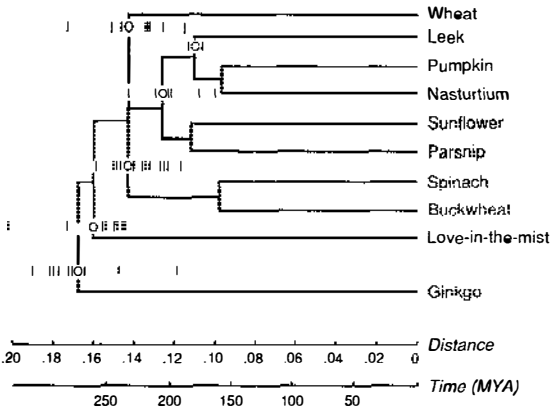


Figure 3 An evolutionary distance tree based on cytochrome-c sequences. Distance is the number of sites that differ between two sequences under comparison divided by the number of sites compared, and MYA is million years ago. The position of each node is the numerical average of distance between the respective families on each side of the node. The isolated dashes designate the specific values upon which this average is based. Wheat and leek represent two monocot families, and pumpkin through love-in-the-mist are from the dicot families cucurbitaceae, brassicaceae, asteracese, capialeae, chenopodinum, polygonacoae, and ranunculi, respectively.

the chloroplast genes and *gapdhA*, also support the notion of an angiosperm radiation more than 200 MYA.

The discussion of the apparent discrepancy between the fossil record and the molecular clock determinations has focused primarily on one of two possibilities. First, it has been suggested that the paleobotanical information is misleading: perhaps a long precretaceous angiosperm history was never captured in the fossil record. This explanation is impossible to formally disprove, but it is considered by those who study plant fossils to be highly unlikely (42). The second possibility is that protein sequence evolution in angiosperms might obey rules different from those of metazoans, and thus be useless for estimating divergence times. That is, the molecular clock does not keep the right time. However, as Martin et al (54) show for *gapdhA*, as Wolfe et al (103) show for the chloroplast sequences, and as we have shown for cytochrome c (93), the rates of angiosperm sequence divergence are comparable to that of the metazoans. In addition, in each case, the relative rate test is satisfied, indicating that the various angiosperm families and outgroup taxa display comparable rates of sequence divergence. These measurements of rate do have error, so that small but significant differences in rates in different lineages under comparison may be missed, but the test is accurate enough to preclude the size of

the discrepancy in divergence times that has emerged between the paleobotanical record and the protein sequence record.

A third explanation—angiosperm evolution driven by horizontal transfer—allows a straightforward interpretation of both the paleobotanical and molecular evidence. This hypothesis easily incorporates the notion that angiosperms are probably polyphyletic. Krassilov has advocated polyphyly (42, 43) to resolve conflicts within the fossil record itself. Polyphyletic origins suggest that modern angiosperms are derived from multiple proangiosperm lineages that first appeared in the early Permian period (ca 260 MYA), and that these multiple lineages underwent concerted conversion into angiosperms in the early Cenozoic era or, as described by Krassilov (42), underwent "angiospermization." It would seem that a theory of angiosperm evolution that incorporates the lateral spread of genes should be taken seriously. To its advantage, it accommodates the otherwise irreconcilable divergence times based on paleobotanical and plant molecular evolution; it resolves conflicts within the fossil record; it answers the appearance of like-traits in distantly related plants that occupy similar ranges [as suggested by Went (99)], and finally it provides a rational explanation for the failure to organize extant angiosperm families into consistent phylogenetic trees.

CONCLUSION

As mentioned in the introduction, geneticists who think at the molecular and cellular level could easily accommodate a theory of evolution incorporating horizontal gene transfer. Other areas of biology would be more seriously affected, though probably not as seriously as feared. Speciation theories that rely on reproductive isolation are perfectly consistent with the notion of macroevolutionary trends that are influenced by horizontal transfers, as illustrated by the examples of *E. coli* and *Salmonella.*

As we have seen with populations of *E. coli* and *S. typhimurium,* horizontal transfers are apparent among subpopulations and with even more distantly related organisms. This has not prevented genetically identifiable subpopulations from being formed, nor has it prevented *Escherichia* and *Salmonella* from being taxonomically associated with the enterics, and the enterics from being identified as members of the purple bacteria. That is, bacteria can be organized into tree-like patterns of descent. As I have previously written (92), it is ironic that after the discovery of transferable plasmids, early workers were criticized for attempting to reconstruct bacterial phylogeny because, it was argued, the horizontal spread of DNA caused by plasmid movement would so scatter traits that tree-like descent patterns would be lost. Some microbiologists studying plasmids had early on assumed cross-species gene transfer, using it to argue against the possibility of phylogenetic classification of bacteria. How-

ever, horizontal gene flow does not necessarily preclude phylogenetic classification but, as we have seen with the angiosperms, it can make it more difficult, just as the existence of reproductive isolation (i.e. the fact that taxonomic barriers do exist) and the fact that these taxonomic groups can be organized into tree-like patterns of descent, does not establish, a priori, the absence of horizontal gene transfer. Evolution favors those groups with an appropriate balance between genetic variation and genetic stability. If genes or gene segments introduced from other species are viewed as just another source of genetic variability, then cross-species gene transfer could logically exist concomitantly with reproductive isolation and other biological barriers that limit its employment and sustenance. Bacteria also provide a model for another aspect of horizontal transfer. In general, the discussion has focused on the transfer of whole genes (see Table 1), whereas the transfer of much shorter elements seems to be common in *E. coli, Streptococci,* and *Neisseria.* I believe that the transfer of parts of genes is also a factor in eukaryotes, as I suggested above for the plant cytochromes c and for mammalian beta-globins (90).

Some neoDarwinian concepts, however, would likely have to be modified if horizontal gene transfer plays an important role in evolution. In particular, the idea that variations that arise within populations are independent of macroevolutionary trends (sometimes referred to as "Wright's" rule) is probably unrealistic, especially with regard to the angiosperms. This notion will likely be replaced with the idea that useful variation, at least in rapidly evolving populations, is influenced by traits carried in competing lineages. This idea was explicitly proposed by Vavilov in 1922 (96) as his "law of homologous variation," before botanists accepted the neoDarwinian paradigm. Vavilov's idea was premature because no mechanism in its support was known. But with the discovery of mechanisms for moving genes horizontally, Went in 1972, and more recently others (42, 43, 47, 71) have suggested that this is possibly the mechanism causing the widespread parallelisms or "homologous variations."

In paleontology, the influence of a theory of macroevolutionary trends influenced by horizontal gene transfers will likely be profound (91). Much of the fossil record is characterized by extensive parallelisms. At the turn of the last century, many paleontologists focused on these observations and constructed theories to accommodate them. These led to "orthogenetic" theories (see ref. 24, for an historical account). With the dominance of neoDarwinian theories, the ideas of orthogenesis have been widely criticized as teleological or even as vitalistic. These criticisms are, for the most part, unwarranted since many orthogenetic theories were offered as legitimate scientific hypotheses. However, because mechanisms to accommodate parallel variations were not known, the theories lapsed. Now, with known mechanisms, these old theories should be reanalyzed in a new light—a process that has already begun by

paleontologists (19, 36). One result of this reanalysis should be the organization of fossils by space and time, much as they were in the past, with less emphasis on identifying uniquely derived characters and on organizing organisms according to cladistic methodology.

Systematic zoological classification will doubtless be influenced by the introduction of horizontal transfers, though the form this may take is difficult to anticipate. There are a number of outstanding problems that current classification schemes have yet to solve. These include, for example, the relationships among the metazoan phyla, the vertebrate classes, and the mammalian orders. From the fossil record, each of these examples arose as a radiation during a brief geological period without clear ancestral-descendant lines. Analysis of both morphological and molecular characters has yet to resolve these branching patterns. Should horizontal gene transfer be recognized as a significant factor by systematists, this will no doubt have a profound influence on future taxonomic theories.

ACKNOWLEDGMENTS

Thanks to my long time collaborator Hy Hartman for many helpful discussions and suggestions, and to Sue Greenwald for ideas and editing.

Literature Cited

1. Arber W. 1991. Elements in microbial evolution. *J. Mol. Evol.* 33:4–12
2. Assali NE, Mache R, Loiseaux-de Goer S. 1990. Evidence for a composite phylogenetic origin of the plastid genome of the brown alga *Pylaiella littoralis* (L. Kjellm). *Plant Mol. Biol.* 15:307–15
3. Assali NE, Martin WF, Sommerville CC, Loiseaux-de Goer S. 1991. Evolution of the Rubisco operon from prokaryotes to algae: structure and analysis of the rbcS gene of the brown alga *Pylaiella littoralis*. *Plant Mol. Biol.* 17: 853–63
4. Baldauf SL, Manhart JR, Palmer JD. 1990. Evolutionary transfer of the chloroplast tufA gene to the nucleus. *Nature* 344:262–65
5. Beltran P, Musser JM, Helmuth R, Farmer JJ III, Frerichs WM, et al. 1988. Toward a population genetic analysis of Salmonella: genetic diversity and relationships among strains of serotypes *S. choleraesuis, S. derby, S. dublin, S. enteritidis, S. heidelberg, S. infantis, S. newport,* and *S. typhimurium*. *Proc. Natl. Acad. Sci. USA* 85:7753–57
6. Bogusz D, Appleby CA, Landsmann J, Dennis ES, Trinick MJ, Peacock WJ. 1988. Functioning haemoglobin genes in non-nodulating plants. *Nature* 331: 178–80
7. Boulter D. 1974. The evolution of plant proteins with special reference to higher plant cytochrome c. *Curr. Adv. Plant Sci.* 4:1–16
8. Boulter D, Gilroy JS. 1992. Partial sequences of 18s ribosomal RNA of 2 genera from each of 6 flowering plant families. *Phytochemistry* 31:1243–46
9. Bowler LD, Zhang QY, Riou JY, Spratt BG. 1994. Interspecies recombination between the penA genes of *Neisseria meningitidis* and commensal *Neisseria* species during the emergence of penicillin resistance in *N. meningitidis*: natural events and laboratory simulation. *J. Bacteriol.* 176:333–37

10. Brennicke A, Grohmann L, Hiesel R, Knoop V, Schuster W. 1993. The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett.* 325:140–55
11. Campbell A, Botstein D. 1983. Evolution of the lambdoid phages. In *LAMBDA II,* ed. RW Hendrix, JW Roberts, FW Stahl, RA Weissberg pp. 365–80. New York: Cold Spring Harbor Lab. Press
12. Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, et al. 1993. Phylogenetics of seed plants—an analysis of nucleotide sequences from the plastid gene rbcl. *Ann. Mo. Bot. Gard.* 80:528–80
13. Clark JB, Maddison WD, Kidwell MG. 1994. Phylogenetic analysis supports horizontal transfer of p transposable elements. *Mol. Biol. Evol.* 11:40–50
14. Degtyarenko KN, Archakov AI. 1993. Molecular evolution of P450 superfamily and P450-containing monooxygenase systems. *FEBS Lett.* 332:1–8
15. Doolittle RF. 1994. Convergent evolution—the need to be explicit. *Trends Biochem. Sci.* 19:15–18
16. Doolittle RF, Feng DF, Anderson KL, Alberro MR. 1990. A naturally occurring horizontal gene transfer from a eukaryote to a prokaryote. *J. Mol. Evol.* 31:383–88
17. Dowson CG, Coffey TJ, Kell C, Whiley RA. 1993. Evolution of penicillin resistance in Streptococcus pneumoniae; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae. Mol. Microbiol.* 9:635–43
18. Doyle JA, Donoghue MJ, Zimmer EA. 1994. Integration of morphological and ribosomal RNA data on the origin of angiosperms. *Ann. Mo. Bot. Gard.* 81: In press
19. Erwin DH, Valentine JW. 1984. "Hopeful monsters", transposons, and metazoan radiation. *Proc. Natl. Acad. Sci. USA* 81:5482–83
20. Fejes E, Masters BS, McCarty DM, Hauswirth WW. 1988. Sequence and transcriptional analysis of a chloroplast insert in the mitochondrial genome of *Zea mays. Curr. Genet.* 13:509–15
20a. Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27: 401–10
21. Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17:368–76
22. Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. 1991. Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *EMBO J.* 10:3073–78
23. Gilbert HJ, Hazlewood GP, Laurie JI, Orpin CG, Xue GP. 1992. Homologous catalytic domains in a rumen fungal xylanase: evidence for gene duplication and prokaryotic origin. *Mol. Microbiol.* 6:2065–72
24. Gould SJ 1977. *Ontogeny and Phylogeny* Cambridge, MA: Harvard Univ. Press/Belknap
25. Gray MW. 1992. The endosymbiont hypothesis revisited. *Int. Rev. Cytol.* 141: 233–357
26. Groisman EA, Saier MH Jr, Ochman H. 1992. Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the Salmonella genome. *EMBO J.* 11:1309–16
27. Groisman EA, Sturmoski MA, Solomon FR, Lin R, Ochman H. 1993. Molecular, functional, and evolutionary analysis of sequences specific to Salmonella. *Proc. Natl. Acad. Sci. USA* 90:1033–37
28. Gruskin KD, Smith TF, Goodman M. 1987. Possible origin of a calmodulin gene that lacks intervening sequences. *Proc. Natl. Acad. Sci. USA* 84:1605–8
29. Guiseppi A, Aymeric JL, Cami B, Barras F, Creuzet N. 1991. Sequence analysis of the cellulase-encoding celY gene of *Erwinia chrysanthemi:* a possible case of interspecies gene transfer. *Gene* 106: 109–14
30. Hartman, H. 1977. Speculation on viruses, cells and evolution. *Evol. Theor.* 3:159–63
31. Hartman H, Syvanen M, Buchanan BB. 1990. Contrasting evolutionary histories of chloroplast thioredoxins f and m. *Mol. Biol. Evol.* 7:247–54
32. Hein J. 1990. Reconstructing evolution of sequences subject to recombination using parsimony. *Math. Biosci.* 98:185–200
33. Hobbs M, Dalrymple BP, Cox PT, Livingstone SP, Delaney SF, Mattick JS. 1991. Organization of the fimbrial gene region of Bacteroides nodosus: class I and class II strains. *Mol. Microbiol.* 5: 543–60
34. Houck MA, Clark JB, Peterson KR, Kidwell MG. 1991. Possible horizontal transfer of Drosophila genes by the mite *Proctolaelaps regalis. Science* 253: 1125–28
35. Hoyer LL, Hamilton AC, Steenbergen SM, Vimr ER. 1992. Cloning, sequencing and distribution of the *Salmonella typhimurium* LT2 sialidase gene, nanH, provides evidence for interspecies gene transfer. *Mol. Microbiol.* 6:873–84

36. Jeppsson L. 1986. A possible mechanism in convergent evolution. *Paleobiology* 12:337–44
37. Kemmerer EC, Lei M, Wu R. 1991. Structure and molecular evolutionary analysis of a plant cytochrome c gene: surprising implications for *Arabidopsis thaliana*. *J. Mol. Evol.* 32:227–37
38. Kidwell M. 1993. Lateral transfer in natural populations of eukaryotes. *Annu. Rev. Genet.* 27:235–56
39. Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol.* 29:170–79
40. Kizawa H, Tomura D, Oda M, Fukamizu A, Hoshino T, et al. 1991. Nucleotide sequence of the unique nitrate/nitrite-inducible cytochrome P-450 cDNA from *Fusarium oxysporum*. *J. Biol. Chem.* 266:10632–37
41. Deleted in proof
42. Krassilov VA. 1977. The origin of angiosperms. *Bot. Rev.* 43:143–76
43. Krassilov VA. 1991. The origin of angiosperms—new and old problems. *Trends Ecol. Evol.* 6:215–20
44. Kraut M, Hugendieck I, Herwig S, Meyer O. 1989. Homology and distribution of CO dehydrogenase structural genes in carboxydotrophic bacteria. *Arch. Microbiol.* 152:335–41
45. Kroll JS, Moxon ER. 1990. Capsulation in distantly related strains of *Haemophilus influenzae* type b: genetic drift and gene transfer at the capsulation locus. *J. Bacteriol.* 172:1374–79
46. Lake JA. 1987. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Mol. Biol. Evol.* 4:1987
47. Lamboy WT. 1984. Evolution of flowering plants by fungus-to-host horizontal gene transfer. *Evol. Theor.* 7:45–51
48. Lee KY, Hopkins JD, Syvanen M. 1991. Evolved neomycin phosphotransferase from an isolate of *Klebsiella pneumoniae*. *Mol. Microbiol.* 5:2039–46
49. Levy SB, Miller RV, eds. 1989. *Gene Transfer in the Environment*. New York: McGraw-Hill
50. Lujan R, Zhang QY, Saez Nieto JA, Jones DM, Spratt BG. 1991. Penicillin-resistant isolates of *Neisseria lactamica* produce altered forms of penicillin-binding protein 2 that arose by interspecies horizontal gene transfer. *Antimicrob. Agents Chemother.* 35:300–4
51. Margulis L. 1971. Symbiosis and evolution. *Sci. Am.* 225:48–57
52. Martin PG, Boulter D, Penny D. 1985. Angiosperm phylogeny studied using sequences of five macromolecules. *Taxon* 34:393–400
53. Martin PG, Dowd JM. 1991. A comparison of 18s ribosomal RNA and rubisco large subunit sequences for studying angiosperm phylogeny. *J. Mol. Evol.* 33: 274–82
54. Martin W, Brinkmann H, Savonna C, Cerff R. 1993. Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. *Proc. Natl. Acad. Sci. USA* 90:8692–6
55. Martin W, Lagrange T, Li YF, Bisanz-Seyer C, Mache R. 1990. Hypothesis for the evolutionary origin of the chloroplast ribosomal protein L21 of spinach. *Curr. Genet.* 18:553–6
56. Martin W, Lydiate D, Brinkmann H, Forkmann G, Saedler H, Cerff R. 1993. Molecular phylogenies in angiosperm evolution. *Mol. Biol. Evol.* 10: 140–62
57. May BP, Dennis PP. 1989. Evolution and regulation of the gene encoding superoxide dismutase from the archaebacterium *Halobacterium cutirubrum*. *J. Biol. Chem.* 264:12253–58
58. Maynard-Smith JM. 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34:126–29
59. Maynard-Smith JM, Dowson CG, Spratt BG. 1991. Localized sex in bacteria. *Nature* 349:29–31
60. Mazodier P, Davies J. 1991. Gene transfer between distantly related bacteria. *Annu. Rev. Genet.* 25:147–71
61. Medigue C, Rouxel T, Vigier P, Henaut A, Danchin A. 1990. Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J. Mol. Biol.* 222:851–6
62. Mereschkowsky C. 1905. Uber Natur und Ursprung der Chromatophoren in Pflanzenteilen. *Biol. Zentrabl.* 25:593–635
63. Milkman R, Bridges MM. 1990. Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. *Genetics* 126:505–17
64. Milkman R, Bridges MM. 1993. Molecular evolution of the *Escherichia coli* chromosome. IV. Sequence comparisons. *Genetics* 133:455–68
65. Mouches C, Bensaadi N, Salvado JC. 1992. Characterization of a LINE retroposon dispersed in the genome of three non-sibling Aedes mosquito species. *Gene* 120:183–90
66. Nakayama S, Kretsinger RH. 1993. Evolution of EF-hand calcium-modulated proteins. III. Exon sequences confirm most dendrograms based on protein sequences: calmodulin dendrograms show

significant lack of parallelism. *J. Mol. Evol.* 36:458–76
67. Ochman H, Selander RK. 1984. Evidence for clonal population structure in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 81:198–201
68. Ohshima K, Koishi R, Matsuo M, Okada N. 1993. Several short interspersed repetitive elements (SINEs) in distant species may have originated from a common ancestral retrovirus: characterization of a squid SINE and a possible mechanism for generation of tRNA-derived retroposons. *Proc. Natl. Acad. Sci. USA* 90:6260–64
69. Palmer JD. 1985. Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* 19:325–54
70. Peacock D, Boulter D. 1975. Use of amino acid sequence data in phylogeny and evaluation of methods using computer simulation. *J. Mol. Biol.* 95:513–27
71. Pirozynski, KA. 1988. Coevolution by horizontal gene transfer: a speculation on the role of fungi. In *Coevolution of Fungi with Plants and Animals*, ed. KA Pirozynski, DL Hawksworth, 11:248–68 London/San Diego: Academic
72. Pritchard AE, Venuti SE, Ghalambor MA, Sable CL, Cummings DJ. 1989. An unusual region of Paramecium mitochondrial DNA containing chloroplast-like genes. *Gene* 78:121–34
73. Radstrom P, Fermer C, Kristiansen BE, Jenkins A, Skold O, Swedberg G. 1992. Transformational exchanges in the dihydropteroate synthase gene of *Neisseria meningitidis:* a novel mechanism for acquisition of sulfonamide resistance. *J. Bacteriol.* 174:6386–93
74. Rayssiguier C, Thaler DS, Radman M. 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* 342:396–401
75. Reeves P. 1993. Evolution of Salmonella O antigen variation hy interspecific gene transfer on a large scale. *Trends Genet.* 9:17–22
76. Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* 6:526–38
77. Selander RK, Smith NH, Li J, Beltran P, Ferris KE, et al. 1992. Molecular evolutionary genetics of the cattle-adapted serovar *Salmonella dublin*. *J. Bacteriol.* 174:3587–92
78. Shatters RG, Kahn ML. 1989. Glutamine synthetase II in Rhizobium: reexamination of the proposed horizontal transfer of DNA from eukaryotes to prokaryotes. *J. Mol. Evol.* 29: 422–28
79. Shay JW, Werbin H. 1992. New evidence for the insertion of mitochondrial DNA into the human genome: significance for cancer and aging. *Mutat. Res.* 275:227–35
80. Shivji MS, Li N, Cattolico RA. 1992. Structure and organization of rhodophyte and chromophyte plastid genomes: implications for the ancestry of plastids. *Mol. Gen. Genet.* 232:65–73
81. Simpson WJ, Musser JM, Cleary PP. 1992. Evidence consistent with horizontal transfer of the gene (emm12) encoding serotype M12 protein between group A and group G pathogenic streptococci. *Infect. Immun.* 60:1890–93
82. Smith MW, Feng DF, Doolittle RF. 1992. Evolution by acquisition: the case for horizontal gene transfers. *Trends Biochem. Sci.* 17:489–93
83. Smith NH, Beltran P, Selander RK. 1990. Recombination of Salmonella phase 1 flagellin genes generates new serovars. *J. Bacteriol.* 172:2209–16
84. Soubrier F, Levy-Schil S, Mayaux JF, Petre D, Arnaud A, Crouzet J. 1992. Cloning and primary structure of the wide-spectrum amidase from *Brevibacterium* sp. R312: high homology to the amiE product from *Pseudomonas aeruginosa*. *Gene* 116:99–104
85. Souillard N, Magot M, Possot O, Sibold L. 1988. Nucleotide sequence of regions homologous to nifH (nitrogenase Fe protein) from the nitrogen-fixing archaebacteria *Methanococcus thermolithotrophicus* and *Methanobacterium ivanovii:* evolutionary implications. *J. Mol. Evol.* 27:65–71
86. Sprague GF Jr. 1991. Genetic exchange between kingdoms. *Curr. Opin. Genet. Dev.* 1:530–33
87. Spratt BG, Bowler LD, Zhang QY, Zhou J, Smith JM. 1992. Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal Neisseria species. *J. Mol. Evol.* 34:115–25
88. Stevens PF. 1984. Metaphors and typology in the development of botanical systematics 1690–1960 or the art of putting new wine in old bottles. *Taxon* 33:169–211
89. Stewart CB, Schilling JW, Wilson AC. 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature* 330:401–4
90. Syvanen M. 1984. Conserved regions in mammalian beta-globins: could they arise by cross-species gene exchange? *J. Theor. Biol.* 107:685–96
91. Syvanen M. 1985. Cross-species gene

transfer; implications for a new theory of evolution. *J. Theor. Biol.* 112:333–43
92. Syvanen M. 1989. Migrant DNA in the microbial world. *Cell* 60:7–8
93. Syvanen M, Hartman H, Stevens PF. 1989. Classical plant taxonomic ambiguities extend to the molecular level. *J. Mol. Evol.* 28: 536–44
94. Troitsky AV, Melekhovets YuF, Rakhimova GM, Bobrova VK, Valiejo-Roman KM, Antonov AS. 1991. Angiosperm origin and early stages of seed plant evolution deduced from rRNA sequence comparisons. *J. Mol. Evol.* 32: 253–61
95. Valentin K, Zetsche K. 1990. Nucleotide sequence of the gene for the large subunit of Rubisco from *Cyanophora paradoxa*—phylogenetic implications. *Curr. Genet.* 18:199–202
96. Vavilov NI. 1922. The law of homologous series in variation. *J. Genet.* 12:47–89
97. Wallin JE. 1927. *Symbioticism and the Origin of the Species.* London: Baillere, Tindall & Cox
98. Wang L, Romana LK, Reeves PR. 1992. Molecular analysis of a *Salmonella enterica* group E1 rfb gene cluster: O antigen and the genetic basis of the major polymorphism. *Genetics* 130: 429–43
99. Went FW. 1971. Parallel evolution. *Taxon* 20:197–226
100. Whatmore AM, Kchoe MA. 1994. Horizontal gene transfer in the evolution of group a streptococcal emm-like genes-gene mosaics and variation in vir regulons. *Mol. Microbiol.* 11:363–74
101. Whittam TS, Ochman H, Selander RK. 1983. Multilocus genetic structure in natural populations of *Escherichia coli. Proc. Natl. Acad. Sci. USA* 80: 1751–55
102. Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* 74:5088–90
103. Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc. Natl. Acad. Sci. USA* 86:6201–5
104. Xiong Y, Eickbush TH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9:3353–62