Mobile Genetic Elements 1:4, 256–261; November/December 2011; © 2011 Landes Bioscience

HGT turbulence

Confounding phylogenetic influence of duplicative horizontal transfer and differential gene conversion

Weilong Hao[1,2,*] and Jeffrey D. Palmer[1]

[1]Department of Biology; Indiana University; Bloomington, IN USA; [2]Department of Biological Sciences; Wayne State University; Detroit, MI USA

Keywords: horizontal gene transfer, gene conversion, maximum likelihood, mitochondria, chimeric gene, phylogeny reconstruction, recombination, simulations

Horizontal gene transfer (HGT) often leads to phylogenetic incongruence. When "duplicative HGT" introduces a second copy of a pre-existing gene, the two copies may then engage in gene conversion, leading to phylogenetically mosiac genes. When duplicative HGT is followed by differential gene conversion among descendant lineages, as under the DH-DC model, phylogenetic analysis is further complicated. To explore the effects of DH-DC on phylogeny reconstruction, we analyzed two sets of sequences: (1) an augmented set of plant mitochondrial *atp1* sequences for which we recently published evidence of DH-DC; and (2) a set of simulated sequences for which we varied the extent of chimerism, the number of chimeric genes and nucleotide substitution rates. We show that the phylogenetic behavior of evolutionarily chimeric genes is highly volatile and depends on both the degree of chimerism and the number of differentially chimeric genes present in the analysis. Furthermore, we show that the presence of chimeric genes in gene trees can spuriously affect the phylogenetic position of purely native sequences, especially by attracting these sequences toward basal positions in trees. We propose the term "HGT turbulence" to describe these complex effects of evolutionarily chimeric genes on phylogenetic results.

Introduction

Horizontal gene transfer (HGT) is very common and of great importance in bacterial evolution[1,2] and is also relatively common in certain eukaryotic lineages.[3,4] One consequence of HGT is incongruence between phylogenies reconstructed from genes with different histories of transfer, including of course no transfer at all. Finding and examining phylogenetically atypical genes has become a standard task in HGT studies.[5-7] In cases where an entire gene has been transferred, without the complication of recombination/conversion with a native homolog, the donor lineage can in principle be identified as the nearest phylogenetic neighbor.[7-9] Gene conversion enters the picture either during the act of gene transfer, when transiently-present foreign DNA directly converts (replaces) part of a native locus[10,11] or, after duplicative HGT, via potentially ongoing gene conversion between co-existing native and foreign copies.[12] Furthermore, gene conversion can occur in either a continuous or discontinuous manner.[13,14] Overall, then, gene conversion can lead to a potentially complex and diverse set of patchwork recombinant sequences, especially if it occurs repeatedly, and differentially, over the course of speciation.[12] Each recombinant gene, if analyzed as a whole, might or might not reflect the true evolutionary history of either or both parental sequences. When parental sequences contribute differentially to the number of informative characters in a recombinant sequence, this sequence will tend to resemble the parental sequence that contributes more informative characters (e.g., refs. 15 and 16), whereas when parental sequences contribute similar numbers of informative characters, the recombinant will potentially be quite different from both parental sequences, depending of course on the degree of divergence of the two parental sequences from each other.[17]

When properly recognized and dealt with, recombination poses few problems for phylogenetic analysis and interpretation. In practice, however, recombination detection is challenging and often subject to failure. First, it is well established that recombination detection programs perform poorly when sequence divergence is low.[18-21] Unfortunately, plant mitochondrial genomes, which collectively constitute a premiere model system for eukaryotic HGT studies,[4,22] usually have very low rates of nucleotide substitution.[23] Second, gene conversion often involves very short tracts of DNA,[13,14] which can make recombination detection very difficult. For instance, ten previously published recombinant regions between plant mitochondrial and chloroplast genes range in length from only 14 to 79 nucleotides.[24,25] Third, existing recombination detection programs are generally designed

*Correspondence to: Weilong Hao; Email: haow@wayne.edu
Submitted: 10/31/11; Revised: 12/11/11; Accepted: 12/13/11
http://dx.doi.org/10.4161/mge.19030

to identify a single or only a small number of recombination breakpoints.[26,27] Intricate gene conversion during the process of duplicative HGT and differential gene conversion (DH-DC) can, however, lead to mosaic gene structures, with multiple foreign regions interspersed with native regions on a fine scale.

We recently reported such mosaicism in mitochondrial *atp1* and *matR* genes belonging to different groups of flowering plants (angiosperms). We show that these mosaic genes largely escaped detection by recombination-detection programs and were recognizable only by direct visual inspection of DNA sequence alignments.[12] In this report, we explore the effects of chimeric sequences on phylogeny reconstruction by conducting phylogenetic analyses on simulated sequences and on an augmented set of the *atp1* sequences analyzed in reference 12.

Results

Phylogenetic analysis of naturally-occurring mosaic genes. In a recent study,[12] we reported the presence of three differentially mosaic types of mitochondrial *atp1* genes in the angiosperm genus Ternstroemia (Pentaphylaceae, Ericales) and concluded that they arose via DH-DC, with the blueberry genus Vaccinium (Ericaceae) the best candidate to be the donor group in the initiating HGT event. As shown in **Figure 1** (adapted from Fig. 2A of ref. 12), each of the three major clades within Ternstroemia possesses a differentially mosaic *atp1* gene, each with multiple (four to five) foreign regions interspersed with native regions. In the one *atp1* phylogeny presented in reference 12, the mosaic Ternstroemia genes were all placed within the Eriaceae, in a paraphyletic relationship with respect to Vaccinium.

To better understand how mosaic genes affect phylogeny reconstruction, we sequenced the mitochondrial *atp1* gene of *Chamaedaphne calyculata*, a close relative of Vaccinium,[28] using the same set of primers and method as in reference 12, and employed it together with varying subsets of previously sequenced Ericales *atp1* genes in maximum likelihood phylogenetic analyses. The *Chamaedaphne atp1* sequence was deposited in GenBank under accession number JN808446. Consistent with our recent study,[12] and in contrast to organismal phylogeny (**Fig. 2A**), in an analysis that included all relevant genes, the three types of mosaic *atp1* genes in Ternstroemia formed a paraphyletic assemblage, with *T. fragrans* sister to the Vaccinium/Chamaedaphne clade, the T-glj clade the most distant from the Vaccinium/Chamaedaphne clade and the T-ips clade in an intermediate position (**Fig. 2B**). Remarkably, when only one type of mosaic *atp1* gene was included in a given analysis, each of the three types fell in a different phylogenetic position (**Fig. 2C–E**). This shows that the phylogenetic position of a mosaic gene can vary depending on the inclusion of additional, related mosaic genes and emphasizes that this position provides little or no reliable information on the nature of the gene's parental sequences.

The only other topological difference among all five *atp1* trees (**Fig. 2B–F**) involves Chamaedaphne, which was weakly placed (41% bootstrap support) within Vaccinium when all mosaic genes were included (**Fig. 2B**), but was placed as sister (with 68–92% support) to a monophyletic Vaccinium in the other four gene trees. This result raises the possibility that the inclusion of mosaic sequences in phylogenetic analyses can affect not only the placement of related mosaic sequences, but also the placement of apparently native sequences.

Phylogenetic analysis of simulated chimeric sequences. We used simulation studies (conducted using Seq-Gen[29]) to further explore the effects of chimeric sequences on phylogeny reconstruction. The following simulation parameters were chosen to be the same as those used in the above analysis of *atp1* sequences: (1) sequence length, 1200 nucleotides; (2) substitution model, GTR; (3) gamma shape parameter, 0.218; (4) proportion of invariant sites: 0.371; (5) nucleotide frequencies, 0.271 (A), 0.207 (C), 0.261 (G) and 0.261 (T); and (6) GTR relative rate parameters, A < - > C = 0.818, A < - > G = 1.938, A < - > T = 0.244, C < - > G = 0.884, C < - > T = 2.219 and G < - > T = 1.000. All but the first of these parameters are based on PhyML[30]

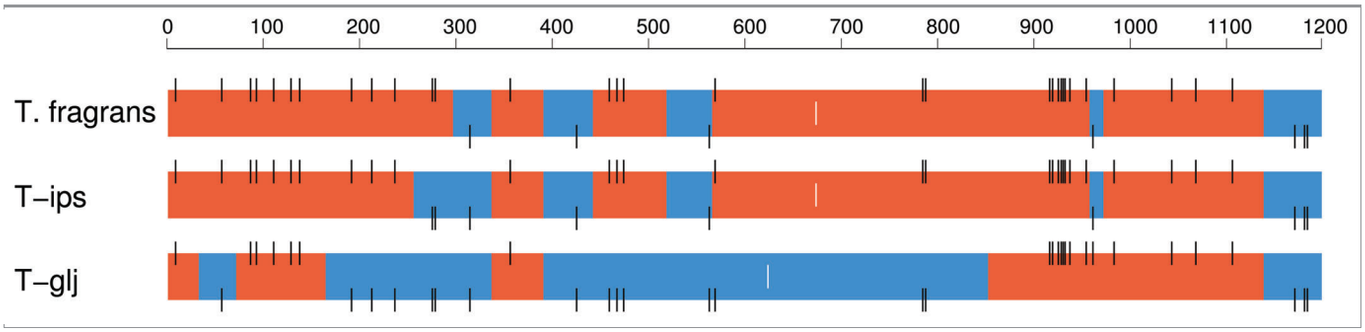

Figure 1. Three types of mosaic mitochondrial *atp1* genes in Ternstroemia (adapted from ref. 12). The multi-colored boxes represent *atp1* genes of the three subclades within Ternstroemia. Black vertical lines represent the 38 nucleotide positions inferred12 to have differed between donor and recipient atp1 genes at the time of atp1 transfer from Vaccinium to a common ancestor of Ternstroemia. Lines at the top of the boxes and red shading indicate sites and regions, respectively, of putatively foreign, Vaccinium ancestry, while bottom lines and blue shading represent native sites and regions. White lines centered within the boxes represent the only two sites that otherwise differ within the Ternstroemia clade. "T-ips" refers to the Ternstroemia subclade containing *T. impressa*, *T. peduncularis* and *T. stahlii*) and "T-glj" to the subclade containing *T. gymnanthera*, *T. longipes* and *T. japonica* (see also **Fig. 2A**).

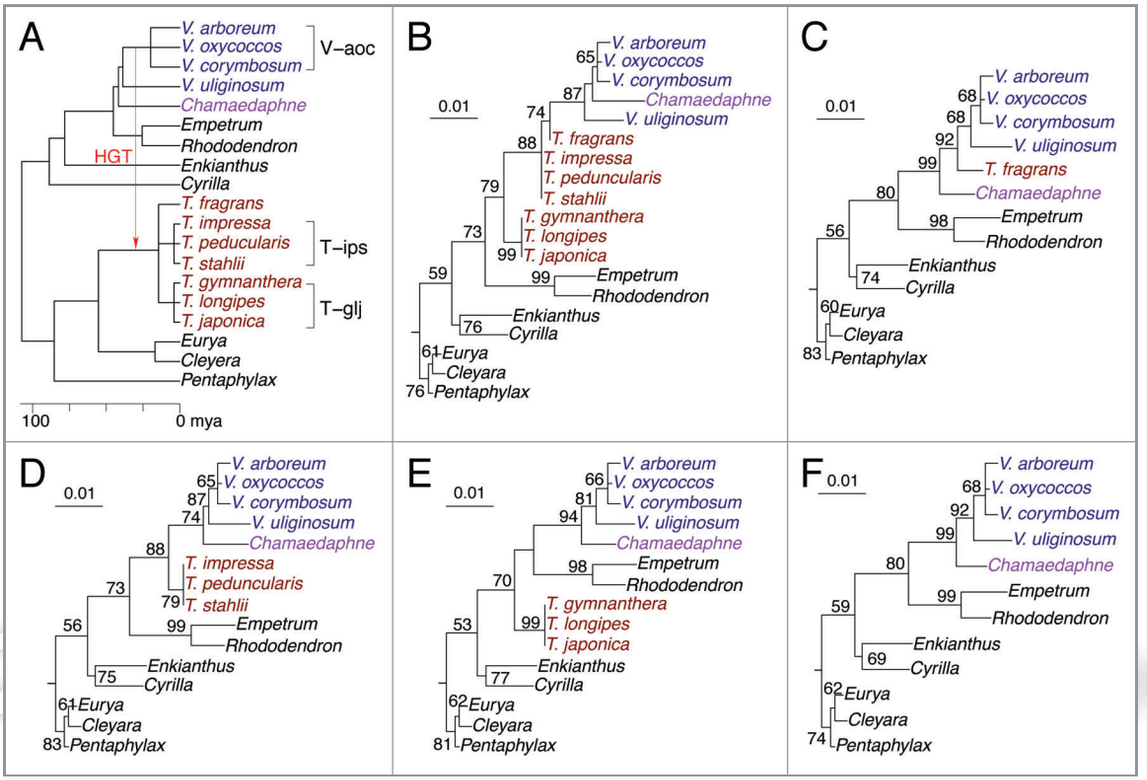

Figure 2. Phylogenetic analysis of mosaic mitochondrial *atp1* genes. (A) Chronogram showing organismal relationships and divergence times of relevant taxa belonging to the Ericales. As described in reference 12, the chronogram was constructed by the BEAST program[43] using a Eurya reference fossil calibration of 86 Myr ago.[44] (B-F) Maximum likelihood phylogenies of mitochondrial *atp1* genes from the taxa shown in (A), with these analyses varying as to which members of Ternstroemia (shown in red), whose *atp1* gene is differentially mosaic, were included. RAxML[32] version 7.0.4 was used to construct all phylogenies with a GTR+Γ+I substitution model. A total of 1000 bootstrap iterations were performed, with all bootstrap values $\geq$ 50% shown on the trees. Phylogenies were rooted using Fouquieria, Marcgravia and Pentamerista as unshown outgroups (hence the stub branch at the base of each gene tree).

estimates derived from the mitochondrial *atp1* alignment shown in **Figure S1**, except with the mosaic Ternstroemia genes excluded (as in **Fig. 2F**) because recombinant genes have been shown to alter the estimation of substitution rate heterogeneity.[31] The remaining simulation parameters were independent of the *atp1* data and include the number of sequences, their topology, relative branch lengths and absolute amount of divergence (**Fig. 3A**). For simplicity, only one recombination breakpoint was allowed in each chimeric sequence, with these chimerics constructed to contain varying proportions (from 50:50 to 10:90) of two of the 16 "parental" sequences generated by the simulations. The 16 purely-simulated sequences together with the artificially constructed chimeric sequence were used in phylogenetic analyses performed using RAxML version 7.0.4.[32]

Two sets of simulation analyses were performed. In the first set (**Fig. 3**), we varied the parental proportions that comprise the chimeric sequence(s) and the number of chimeric sequences included in a given analysis. In analyses with a single chimeric sequence, this sequence grouped with 100% bootstrap support with its majority parental sequence when the parental ratio was 10:90 (i.e., when the chimeric sequence consisted of 10% of sequence 9 and 90% of sequence 1; **Fig. 3B**). The same topology was obtained with the 30:70 chimera (**Fig. 3C**), but the bootstrap value dropped to 92. When the chimera was 50:50 (**Fig. 3D**), it went to the base of the tree, in between the two main clades of simulated sequences, and with bootstrap support reduced along the branches leading to both parental sequences. These results thus show that the phylogenetic position of chimeric genes can vary substantially as the proportion of parental sequences that comprise the chimerics varies.

The phylogenetic position of the 50:50 chimeric sequence also varied substantially depending on whether it was the only chimeric sequence in the analysis (**Fig. 3D**) or whether the 30:70 and/or 10:90 sequences were also included (**Fig. 3E–G**). Inclusion of the 30:70 sequence, either with (**Fig. 3E**) or without (**Fig. 3F**) the 10:90, "pulled" the 50:50 from the base of the tree to its periphery, together with the 30:70 and parental sequence 1 (and with the 10:90 when included) and with strong support (96% and 92%, respectively). This peripheral attraction was more subdued when the 50:50 was paired with the 10:90 (**Fig. 3G**) as opposed to the 30:70 (**Fig. 3F**), presumably because of the greater proportion of sequence length shared by the 30:70 and 50:50 (80%) relative to the 10:90 and 50:50 (60%). Also, there is evidence for a mutual attraction between the 10:90 and 50:50 (but not between the 30:70 and

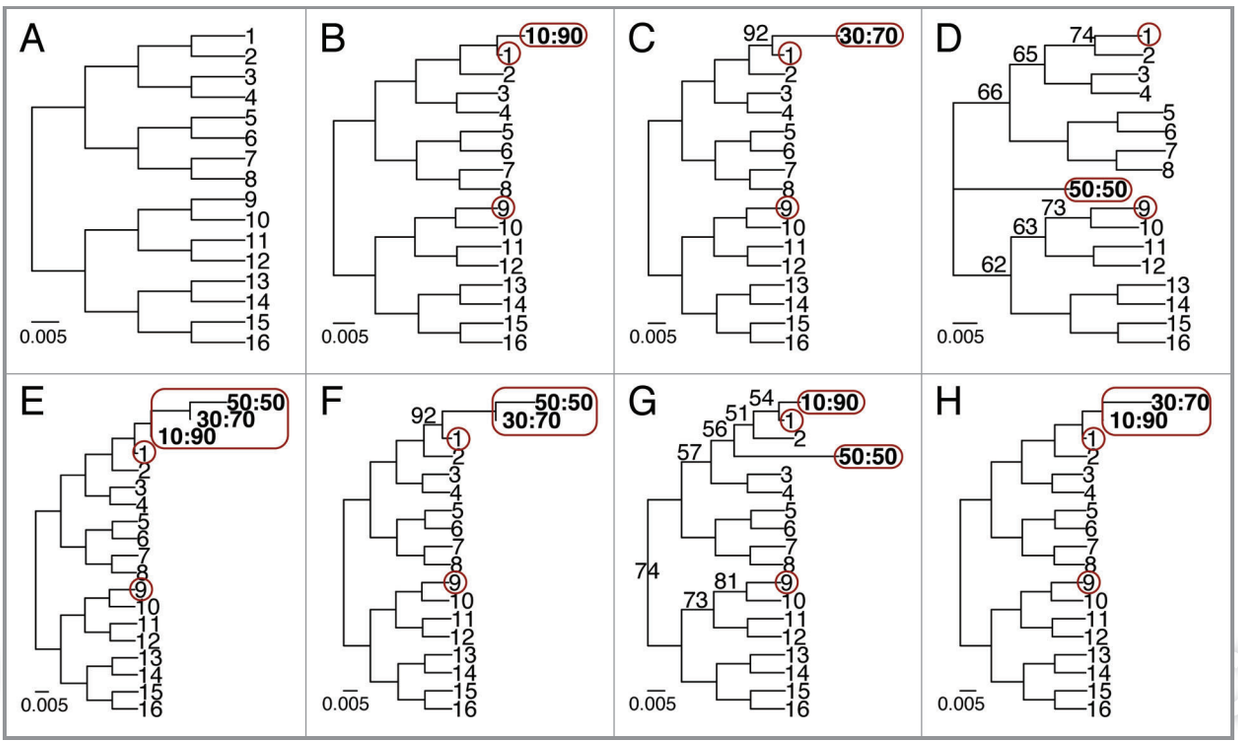

Figure 3. Variable phylogenetic placement of chimeric genes demonstrated by simulations. Artificial sequences were simulated as described in the text. Chimeric sequences were generated by combining the 5′ and 3′ portions of sequences 1 and 9, respectively (both circled in red), with different length ratios (10:90, 30:70 or 50:50) of the two parental sequences. One thousand tree-building iterations were performed. The tree shown in each panel is based (for computational ease) on the concatenated sequences of the first 100 iterations, while the bootstrap support values are from all 1000 trees. All bootstrap values are ≥ 95% except for those shown.

50:50), in that the bootstrap support for the 10:90 grouping with parental sequence 1 was reduced from 100% (**Fig. 3B**) to only 54% (**Fig. 3G**). These simulation results are consistent with the empirical results (**Fig. 2**) in showing that the phylogenetic position of a chimeric gene can vary substantially when different additional chimeric genes are sampled.

We also found that chimeric genes can perturb phylogenetic analysis by producing branch length bias. First, certain chimeric genes were directly associated with elongated branch lengths. In the single-chimeric analyses, the 10:90 chimera had a notably longer branch length than sequence 1 (**Fig. 3B**), while the 30:70 branch length was even longer (**Fig. 3C**), with the (homoplasious) substitutions contributed by the 30% of the chimeric sequence originating from sequence 9 presumably responsible for this greatly extended branch length. Similar results were obtained in **Figure 3E–H**, where multiple chimeric sequences were included in each analysis. Second, the presence of the 50:50 chimeric in the single-chimeric analysis of **Figure 3D** resulted in the loss of a molecular clock among the non-recombinant sequences in this analysis, with the branch leading to the (1–4) clade only 60% of the length leading to the (5–8) clade and the branch leading to the (9–12) clade only 58% of the length leading to the (13–16) clade. There are consistently two additional—albeit much less pronounced—sets of branch length differences in the six trees (**Fig. 3B, C and E–H**) in which one or more chimeric sequences are sister to sequence 1.

The second set of simulations showed that chimeric sequences can also confound phylogenetic analysis by altering the placement of non-recombinant (native) sequences. In these simulations (**Fig. 4**), we varied the absolute amount of sequence divergence across the tree, with the 50:50 sequence the only chimeric sequence in each analysis. **Figure 4A** shows the same tree as **Figure 3D**, while **Figure 4B–D** has expanded divergence by a factor of 2 ×, 5 × and 10 × , respectively. As the simulated sequences become more divergent, the parental sequences show an increased and pronounced, tendency to be attracted toward the base of the tree by the chimeric gene. Also, the branch leading to the chimeric sequence becomes increasingly short, approaching zero in **Figure 4C and D**. It is important to note that this effect is not simply the result of there being increasingly more informative characters from **Figure 4A–D**. In analyses with 5 × or 10 × divergence (data not shown), but only 1/10th the sequence length (120 nucleotides rather than the 1200 in the simulations shown in **Figure 4**), the topology and bootstrap values were essentially identical with those in **Figure 4C and D**, respectively, with these being substantially different from those in **Figure 4A and B**. The different topologies shown in **Figure 4** must therefore result from deterministic effects arising from the varying levels of divergence in these trees. These simulations thus show that inclusion of chimeric sequences can distort the branching pattern of nonchimeric,

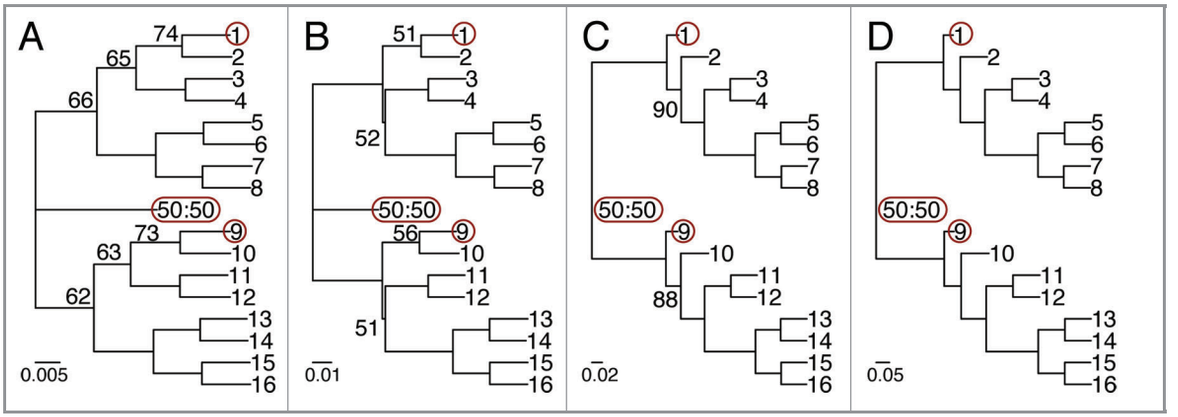

Figure 4. Chimeric sequences increasingly disrupt native-sequence topologies as divergence increases. (A) The same tree as shown in **Figure 3D**. (B–D) Maximum likelihood analysis of the same simulated sequences, but with increasing branch-length scales across the set of sequences: (B) 2 × the scale in (A); (C) 5 × ; and (D) 10 × . One thousand tree-building iterations were performed. The tree shown in each panel is based (for computational ease) on the concatenated sequences of the first 100 iterations, while the bootstrap support values are from all 1,000 trees. All bootstrap values are ≥ 95% except for those shown.

native sequences and that the extent of this distortion varies directly with the absolute level of sequence divergence.

Discussion

The analyses reported in this study show that the inclusion in phylogenetic analyses of chimeric sequences arising from HGT and gene conversion can produce a variety of spurious phylogenetic results. These include the misplacement of both chimeric and native sequences, as well as branch length distortions. Accordingly, we introduce the term "HGT turbulence" as a general moniker for this category of phylogenetic artifacts. Mutually reinforcing evidence for these types of HGT turbulence was apparent in the phylogenetic analyses of both simulated sequences and a naturally occurring set of native and chimeric mitochondrial sequences, with the simulations providing greater opportunity to crystallize and illustrate specific sets of sequence interactions and consequences. It is also important to realize that while these simulations were framed and presented in the context of HGT, their results apply equally well to conversion between paralogs arising from internal gene duplication as to xenologs arising from HGT.

HGT turbulence is probably relatively common in bacteria,[1,2] given the prevalence of HGT and recombination during bacterial genome evolution. For example, such diverse bacteria as *Neisseria meningitidis*,[33,34] *Streptococcus pneumoniae*,[35,36] *Helicobacter pylori*[37] and Wolbachia[38] have been found to be so recombinogenic that scientists have resorted to using multiple loci (e.g., multilocus sequence typing[39]) as opposed to a single locus to identify clones. Surprisingly, however, the phenomenon of HGT turbulence has never been explicitly addressed in the bacterial literature. One reason for this is that these studies have mainly focused on minimizing the effect of recombination and thereby inferring accurate evolutionary relationships, or on quantifying the number of recombinant genes, rather than actually exploring the topological alterations caused by HGT turbulence. Also, the precise origins of horizontally transferred genes (or gene fragments) in bacterial genomes can be extremely difficult to recover, especially when transfer and/or subsequent recombination have occurred on a fine scale.

Several other studies[17,31,40-42] have used recombinant sequences in phylogenetic simulations, but most of these have focused on the issue of whether recombinants are detectable and/or how to detect them. Perhaps most relevant to our study is the 2002 study by Posada and Crandall,[17] which also explored the relationship between the location of recombination breakpoints and the phylogenetic placement of recombinant sequences, reaching similar conclusion to ours on this point. However, none of these studies explored the effects of combining multiple, related chimeric sequences in the same analysis, nor did they show that chimeric sequences can, under certain conditions, substantially alter the phylogenetic behavior of native sequences. Further simulation studies on the effects of recombination on phylogenetic inference would therefore appear to be called for.

Recognition of the remarkable frequency and extent of horizontal gene transfer, and its often great evolutionary importance, is arguably the greatest accomplishment of the past 15 years of comparative genomics research. Because HGT is so common and important, recognizing and properly dealing with HGT turbulence is likewise important. This is so not only because of the obvious need for obtaining accurate estimates of gene and species phylogeny, but also because otherwise too many cases of chimeric HGT, including complex situations involving DH-DC, will continue to go overlooked.

Acknowledgments

We thank the two reviewers for critical reading of the manuscript; the Kew DNA Bank (Royal Botanic Gardens; Kew, UK) for providing the *Chamaedaphne calyculata* DNA and Y. Zheng for technical assistance. W.H. was supported by startup funds from Wayne State University. J.D.P. was supported by National Institutes of Health Grant R01-GM-70612 and by the METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc.

Supplemental Materials

Supplemental materials can be found at: www.landesbioscience.com/journals/mge/article/19030

References

1. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature 2000; 405:299-304; PMID:10830951; http://dx.doi.org/10.1038/35012500
2. Beiko RG, Harlow TJ, Ragan MA. Highways of gene sharing in prokaryotes. Proc Natl Acad Sci USA 2005; 102:14332-7; PMID:16176988; http://dx.doi.org/10.1073/pnas.0504068102
3. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 2008; 9:605-18; PMID:18591983; http://dx.doi.org/10.1038/nrg2386
4. Bock R. The give-and-take of DNA: Horizontal gene transfer in plants. Trends Plant Sci 2010; 15:11-22; PMID:19910236; http://dx.doi.org/10.1016/j.tplants.2009.10.001
5. Shi T, Falkowski PG. Genome evolution in cyanobacteria: the stable core and the variable shell. Proc Natl Acad Sci USA 2008; 105:2510-5; PMID:18268351; http://dx.doi.org/10.1073/pnas.0711165105
6. Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. Genome Res 2006; 16:1099-108; PMID:16899658; http://dx.doi.org/10.1101/gr.5322306
7. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 2003; 424:197-201; PMID:12853958; http://dx.doi.org/10.1038/nature01743
8. Mower JP, Stefanovic S, Young GJ, Palmer JD. Plant genetics: gene transfer from parasitic to host plants. Nature 2004; 432:165-6; PMID:15538356; http://dx.doi.org/10.1038/432165b
9. Moran NA, Jarvik T. Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science 2010; 328:624-7; PMID:20431015; http://dx.doi.org/10.1126/science.1187113
10. Chan CX, Beiko RG, Darling AE, Ragan MA. Lateral transfer of genes and gene fragments in prokaryotes. Genome Biol Evol 2009; 1:429-38; PMID:20333212; http://dx.doi.org/10.1093/gbe/evp044
11. Keeling PJ, Palmer JD. Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. Proc Natl Acad Sci USA 2001; 98:10745-50; PMID:11526220; http://dx.doi.org/10.1073/pnas.191337098
12. Hao W, Richardson AO, Zheng Y, Palmer JD. Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. Proc Natl Acad Sci USA 2010; 107:21576-81; PMID:21115831; http://dx.doi.org/10.1073/pnas.1016295107
13. Palmer S, Schildkraut E, Lazarin R, Nguyen J, Nickoloff JA. Gene conversion tracts in *Saccharomyces cerevisiae* can be extremely short and highly directional. Nucleic Acids Res 2003; 31:1164-73; PMID:12582235; http://dx.doi.org/10.1093/nar/gkg219
14. Schaeffer SW, Anderson WW. Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. Genetics 2005; 171:1729-39; PMID:16143625; http://dx.doi.org/10.1534/genetics.105.041947
15. Mower JP, Stefanovic S, Hao W, Gummow JS, Jain K, Ahmed D, et al. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. BMC Biol 2010; 8:150; PMID:21176201; http://dx.doi.org/10.1186/1741-7007-8-150
16. Schönenberger J, Anderberg AA, Sytsma KJ. Molecular phylogenetics and patterns of floral evolution in the Ericales. Int J Plant Sci 2005; 166:265-88; http://dx.doi.org/10.1086/427198
17. Posada D, Crandall KA. The effect of recombination on the accuracy of phylogeny estimation. J Mol Evol 2002; 54:396-402; PMID:11847565
18. Drouin G, Prat F, Ell M, Clarke GD. Detecting and characterizing gene conversions between multigene family members. Mol Biol Evol 1999; 16:1369-90; PMID:10563017
19. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci USA 2001; 98:13757-62; PMID:11717435; http://dx.doi.org/10.1073/pnas.241370698
20. Wiuf C, Christensen T, Hein J. A simulation study of the reliability of recombination detection methods. Mol Biol Evol 2001; 18:1929-39; PMID:11557798
21. Hao W. OrgConv: Detection of gene conversion using consensus sequences and its application in plant mitochondrial and chloroplast homologs. BMC Bioinformatics 2010; 11:114; PMID:20196863; http://dx.doi.org/10.1186/1471-2105-11-114
22. Richardson AO, Palmer JD. Horizontal gene transfer in plants. J Exp Bot 2006; 58:1-9; PMID:17030541; http://dx.doi.org/10.1093/jxb/erl148
23. Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol 2007; 7:135; PMID:17688696; http://dx.doi.org/10.1186/1471-2148-7-135
24. Hao W, Palmer JD. Fine-scale mergers of chloroplast and mitochondrial genes create functional, trans-compartmentally chimeric mitochondrial genes. Proc Natl Acad Sci USA 2009; 106:16728-33; PMID:19805364; http://dx.doi.org/10.1073/pnas.0908766106
25. Sloan DB, Alverson AJ, Storchova H, Palmer JD, Taylor DR. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. BMC Evol Biol 2010; 10:274; PMID:20831793; http://dx.doi.org/10.1186/1471-2148-10-274
26. Martin DP. Recombination detection and analysis using RDP3. Methods Mol Biol 2009; 537:185-205; PMID:19378145; http://dx.doi.org/10.1007/978-1-59745-251-9_9
27. Husmeier D, Wright F, Milne I. Detecting interspecific recombination with a pruned probabilistic divergence measure. Bioinformatics 2004; 20:1806-7; PMID:14988107
28. Kron KA, Judd WS, Crayn DM. Phylogenetic analyses of Andromedeae (Ericaceae subfam. Vaccinioideae). Am J Bot 1999; 86:1290-300; PMID:10487817; http://dx.doi.org/10.2307/2656777
29. Rambaut A, Grassly NC. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput Appl Biosci 1997; 13:235-8; PMID:9183526
30. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010; 59:307-21; PMID:20525638; http://dx.doi.org/10.1093/sysbio/syq010
31. Schierup MH, Hein J. Consequences of recombination on traditional phylogenetic analysis. Genetics 2000; 156:879-91; PMID:11014833
32. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006; 22:2688-90; PMID:16928733; http://dx.doi.org/10.1093/bioinformatics/btl446
33. Feil EJ, Maiden MC, Achtman M, Spratt BG. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. Mol Biol Evol 1999; 16:1496-502; PMID:10555280
34. Hao W, Ma JH, Warren K, Tsang RS, Low DE, Jamieson FB, et al. Extensive genomic variation within clonal complexes of *Neisseria meningitidis*. Genome Biol Evol 2011; 3:1406-18; PMID:22084315; http://dx.doi.org/10.1093/gbe/evr119
35. Lefébure T, Stanhope MJ. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol 2007; 8:R71; PMID:17475002; http://dx.doi.org/10.1186/gb-2007-8-5-r71
36. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science 2011; 331:430-4; PMID:21273480; http://dx.doi.org/10.1126/science.1198545
37. Saunders NJ, Boonmee P, Peden JF, Jarvis SA. Inter-species horizontal transfer resulting in core-genome and niche-adaptive variation within *Helicobacter pylori*. BMC Genomics 2005; 6:9; PMID:15676066; http://dx.doi.org/10.1186/1471-2164-6-9
38. Klasson L, Westberg J, Sapountzis P, Näslund K, Lutnaes Y, Darby AC, et al. The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. Proc Natl Acad Sci USA 2009; 106:5725-30; PMID:19307581; http://dx.doi.org/10.1073/pnas.0810753106
39. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA 1998; 95:3140-5; PMID:9501229; http://dx.doi.org/10.1073/pnas.95.6.3140
40. Beiko RG, Doolittle WF, Charlebois RL. The impact of reticulate evolution on genome phylogeny. Syst Biol 2008; 57:844-56; PMID:19085328; http://dx.doi.org/10.1080/10635150802559265
41. Arenas M, Posada D. The effect of recombination on the reconstruction of ancestral sequences. Genetics 2010; 184:1133-9; PMID:20124027; http://dx.doi.org/10.1534/genetics.109.113423
42. Ruths D, Nakhleh L. Recombination and phylogeny: effects and detection. Int J Bioinform Res Appl 2005; 1:202-12; PMID:18048130; http://dx.doi.org/10.1504/IJBRA.2005.007578
43. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 2007; 7:214; PMID:17996036; http://dx.doi.org/10.1186/1471-2148-7-214
44. Bremer K, Friis EM, Bremer B. Molecular phylogenetic dating of asterid flowering plants shows early Cretaceous diversification. Syst Biol 2004; 53:496-505; PMID:15503676; http://dx.doi.org/10.1080/10635150490445913