J Mol Evol (2014) 78:163–170
DOI 10.1007/s00239-013-9607-7

ORIGINAL ARTICLE

Quartet Analysis of Putative Horizontal Gene Transfer in Crenarchaeota

Travers H. Ching · Brandon A. Yoza · Qing X. Li

Received: 13 June 2013 / Accepted: 6 December 2013 / Published online: 18 December 2013

Abstract Horizontal gene transfers (HGT) between four Crenarchaeota species (*Metallosphaera cuprina* Ar-4^T, *Acidianus hospitalis* W1^T, *Vulcanisaeta moutnovskia* 768-28^T, and *Pyrobaculum islandicum* DSM 4184^T) were investigated with quartet analysis. Strong support was found for individual genes that disagree with the phylogeny of the majority, implying genomic mosaicism. One such gene, a ferredoxin-related gene, was investigated further and incorporated into a larger phylogeny, which provided evidence for HGT of this gene from the *Vulcanisaeta* lineage to the *Acidianus* lineage. This is the first application of quartet analysis of HGT for the phylum Crenarchaeota. The results have shown that quartet analysis is a powerful technique to screen homologous sequences for putative HGTs and is useful in visually describing genomic mosaicism and HGT within four taxa.

Keywords Quartet analysis · Horizontal gene transfer · Crenarchaeota · Genetic mosaicism

Introduction

Quartet analysis was first described by Strimmer and Von Haeseler (1997) utilizing posterior probabilities to analyze the phylogenic relationships among four closely related species. They used simulations to differentiate two different modes of evolution: star phylogeny, in which four taxa evolved simultaneously (i.e., a polytomy) versus bifurcation, in which taxa came about by splitting lineages in two. Strimmer and Von Haeseler (1997) demonstrated that quartet analysis is a powerful technique that provides information that standard phylogenetic techniques could not. Within a genera, the computational creation of standard phylogenies for the comparison of all gene pairs, even with relatively small genome sizes are prohibitive. Furthermore, quartet analysis can also be useful as a technique for the verification of standard rDNA sequencing and phylogenic reconstruction, insuring that the majority of other genes support specific relationships in rDNA trees.

Since the original paper by Strimmer and Von Haeseler, other authors have utilized quartet analysis to demonstrate genomic mosaicism and horizontal gene transfer (HGT) by observing phylogenetic discrepancies in different genes within the same group (Zhaxybayeva and Gogarten 2002, 2003). This technique has also been improved by using more conservative branch support measures such as bootstrap values or by including additional taxa in tree reconstructions before mapping to reduce the influence of long-branch attraction (Zhaxybayeva and Gogarten 2002, 2003).

Crenarchaeota is a phylum in Archaea well known for its extreme thermophilic members, although recent studies have shown them to be ubiquitous in the marine environment (DeLong 1992, Fuhrman 1992, Barns et al. 1996). Crenarchaeota are also known to have high levels of HGT (Barns et al. 1996; Ribeiro and Golding 1998; Young 2001;

Electronic supplementary material The online version of this article (doi:10.1007/s00239-013-9607-7) contains supplementary material, which is available to authorized users.

T. H. Ching · B. A. Yoza (✉)
Department of Microbiology, University of Hawaii at Manoa, Honolulu, HI 96822, USA
e-mail: byoza@hawaii.edu

Q. X. Li
Department of Molecular Biosciences and Bioengineering, University of Hawaii at Manoa, Honolulu, HI 96822, USA

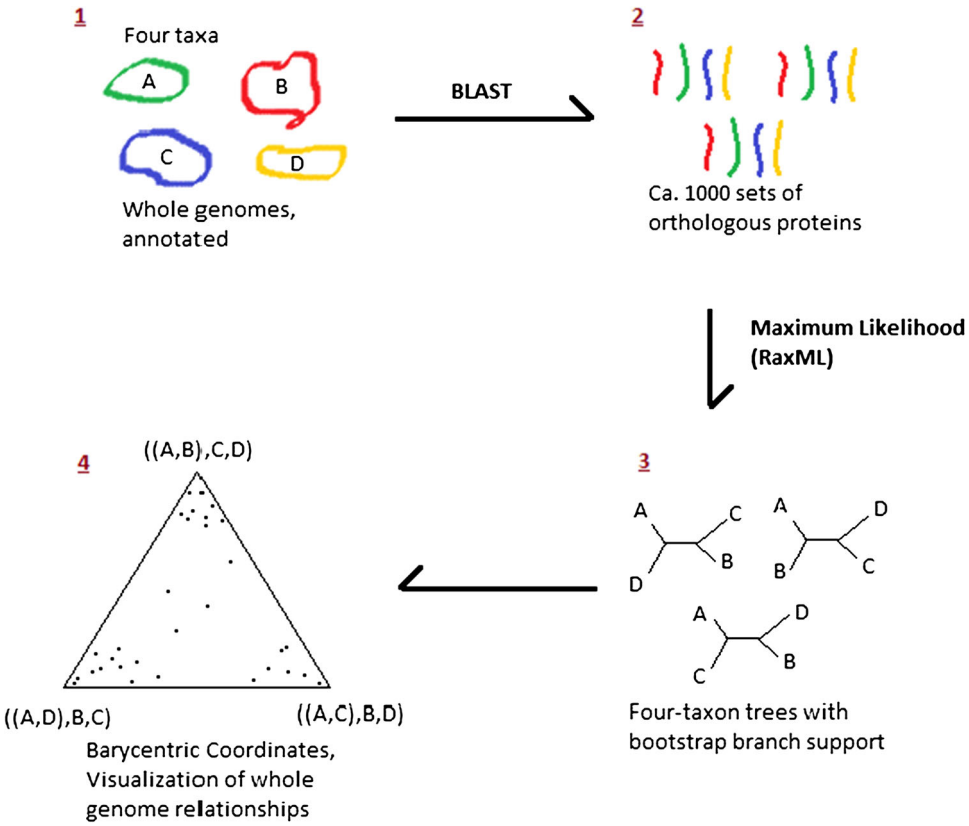

Fig. 1 Workflow of quartet analysis (details are described within the text). Four taxa were chosen (1.1); QuartOPs were found using BLAST (1.2); QuartOPs were bootstrapped and were used to create maximum likelihood phylogenies (1.3) and were then graphed in barycentric coordinates (1.4)

Kunin et al. 2005; Reno et al. 2009; Nelson et al. 1999). However, a comprehensive analysis of HGT within the phylum has not been performed; there is a need to develop methods that are comprehensive, quantitative, and statistical (Ragan 2001).

In the literature, putative HGTs provide only indeterminate evidence when they are not put into a phylogenetic or systematic framework. Reciprocal BLAST and other distance based methods (Reno et al. 2009; Garcia-Vallvé et al. 2000) are not sufficient to eliminate other possibilities, such as gene loss and long-branch attraction. Garcia-Vallvé et al. (2000), therefore, suggested that their findings should represent a "first approximation" of the extent of HGT within the taxa studied. In their critical review of HGT studies, Kurland et al. (2003) discussed how distance based methods can lead to false positives. They suggested that the widespread use of BLAST as a means of determining HGT has systematically inflated the extent and importance of HGT. Salzberg et al. (2001) demonstrated just how dangerous distance based methods can be when providing erroneous results. They reviewed published data that purported HGT from bacteria to human based on protein BLAST searches between genomes of different organisms. They concluded that the results were premature and demonstrated that gene loss and consequently long-branch attraction were much more likely the cause of the genomic mosaicism that linked human and bacteria genes.

The workflow of quartet analysis in Fig. 1 is illustrated as a means to analyze genomic mosaicism and HGT. A quartet of orthologous proteins (QuartOP) from four taxa has only three possible topologies with one branch support value per topology, and can be graphed in two dimensions on an equilateral triangle called barycentric coordinates (Strimmer and Von Haeseler 1997; Zhaxybayeva and Gogarten 2002, 2003). Each corner of the triangle is mutually exclusive and represents one possible tree topology. The closer a point is to a corner, the more support it lends to that topology. A set of four taxa that has strongly supported relationships will have most or all of its data points within one particular corner.

At the time of analysis, 35 annotated Crenarchaeota genomes were downloaded from the UCSC Archaea Genome Database (archaea.ucsc.edu). It was hypothesized that this workflow could be used to distinctly demonstrate HGT within these 35 genomes without a priori knowledge of biogeography, molecular physiology, or other such information. HGT can be determined through purely analytical approaches.

Four taxa were chosen for quartet analysis. The results suggested horizontal transfer of a ferredoxin-related gene. At the end of this analysis, we discuss how the workflow could be used to analyze intra-phylum HGT more comprehensively. We also discuss the strengths and limitations

of this method. The four taxa chosen were: *Metallosphaera cuprina* Ar-4^T, *Acidianus hospitalis* W1^T, *Vulcanisaeta moutnovskia* 768-28^T, and *Pyrobaculum islandicum* DSM 4184^T.

Methods

Preliminary Analysis of 16S/26S Phylogeny

From the UCSC Archaea Genome Database (archaea.ucsc.edu), 35 annotated Crenarchaeota genomes were downloaded. A distant outgroup within a different archaeal phylum, *Archaeoglobus profundus* DSM 5631^T, was also used for analysis. Using a Perl script, 16S and 26S sequences were extracted from the genome annotations and assembled into a fasta file, which was subsequently aligned with ClustalW using standard settings (Gap Opening Penalty: 15, Gap Extension Penalty: 6.66, Weight Matrix: IUB). These alignments were used to build neighbor joining and maximum likelihood trees (Supplemental files A) in MEGA 5.05 (for maximum likelihood: general time reversible model G+I, partial deletion of gaps, nearest neighbor interchange for heuristic search, and 500 bootstraps; for neighbor-joining: Tamura-Nei model, partial deletion of gaps, and 500 bootstraps) (Tamura et al. 2011).

Choice of Quartet

When selecting the four species for quartet analysis, the following need to be considered for selection of the two sets of two paired species: (1) each paired species should have a short branch length between them in the 16S phylogeny in order to reduce the chances for long-branch attractions and (2) the distance between the two separate pairs is large relative to the branch lengths between species in each pair, so that the QuartOP topologies as well as 16S topology would be readily apparent and clearly supported. The two pairs were *M. cuprina* Ar-4^T and *A. hospitalis* W1^T (with a branch distance between them of 0.09 in the maximum likelihood phylogeny; *V. moutnovskia* 768-28^T, and *P. islandicum* DSM 4184^T with a branch distance of 0.06. The major node separating the two groups had a strong bootstrap branch support of 95 and the branch distance between the two groups was 0.112.

Generating QuartOPs

The annotated protein sequences from each species in the quartet were then extracted using BioPerl's searchIO function, and these sequences were used to create a BLAST database which included c.a. 1,000 proteins per genome using the makeblastdb command from NCBI's command line BLAST suite. The size of the database for BLASTP was determined by the total described proteins in the database for the compared species. The QuartOPs were determined with the BLASTP function using the coding sequences from one of the quartet species, in this examination the *M. cuprina* genome as query sequences. The query species used should be selected based upon the largest available genome. There are 243 ORFs in the *M. cuprina* genome that do not occur in its closest neighbor, *Metallosphaera sedula* (Liu et al. 2011). The other species of the quartet were also used as queries and no significant differences were observed. The expect value (E) was set at 1e-6 (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ#expect). The lower expect value used results in a better statistically significant match. These generated top hits (highest bit score) from each of the compared sequences together with the query sequence from *M. cuprina* formed individual QuartOPs, that were placed into a fasta file and used as the quartets in the maximum likelihood reconstruction. A considerable reduction in quartOPs was observed during our examination, when using bidirectional BLAST. For the current analysis a bidirectional BLAST was not performed as it results in the potential exclusion of true orthologs (Wolf and Koonin 2012).

Alignment, Tree Building, and Graphing

Each QuartOP was aligned with ClustalW using default settings and used to build bootstraps of four species maximum likelihood trees using RAxML (matrix: PROTGAMMABLOSUM62F, bootstraps: 500) (Stamatakis 2006). The generated bootstrap support values were then graphed in barycentric coordinates using a modified excel template developed by W. Vaughan (wvaughan.org/ternaryplots.html). These processes were automated using a Perl script to transfer output from one program to the next.

Genomic Mosaicism: HGT Versus Gene Loss

From eight putative genes with bootstrap values greater than 95 and observed in the barycentric plot (Fig. 2, corner B), one QuartOP (a ferredoxin iron-sulfur binding domain protein) disagreed strongly with the majority topology having the greatest bootstrap support and was chosen for further analysis. The values in the barycentric plot are mutually exclusive as a quartOP cannot occupy more than one corner. The other 7 putative genes were not analyzed in this study. Orthologs from 35 genomes were found by blastp, and a maximum likelihood tree was built using MEGA. Compared with RAxML, MEGA provides better graphic support and was useful for confirmation of RAxML results. This tree was then compared with the maximum likelihood 16S phylogeny.

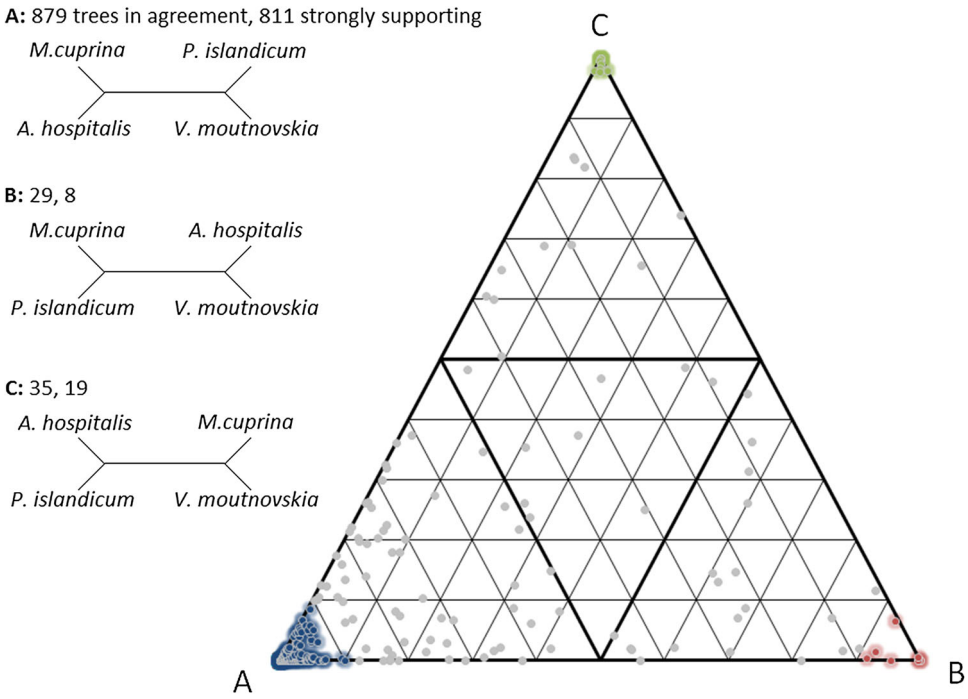

Fig. 2 Plot of branch support values in barycentric coordinates of the chosen quartet. The topology (*A*) had 879 QuartOPs that agreed and 811 that are strongly supported (>95 bootstrap). (*B*) Had 29 QuartOPs that agreed and 8 that are strongly supported. (*C*) Had 35 QuartOPs that agreed and 19 that are strongly supported. This topology for (*A*) was in opposition with the 93 % majority. The QuartOP for (*B*) was chosen for further analysis as it strongly supported the topology of (*A. hospitalis, V. moutnovskia*) and (*M. cuprina, P. islandicum*)

Results

16S and 26S Sequence Analysis for Quartet Selection

When performing sequence analysis for the selection of quartets, the phylogenies with the best basal branch support should be utilized for comparison. In comparison with 16S, the 26S phylogeny had less branch support at the basal branches in both maximum likelihood and neighbor joining trees, yet did not improve branch support at the apical branches (Supplemental file A). The outgroup also fell between Crenarchaeota species in both 26S phylogenies indicating long-branch attraction. The 16S maximum likelihood and neighbor joining phylogenies produced very similar topologies, the basal outgroup branched earlier than other members within the phylum. The 16S phylogenies had a better basal branch support in comparison with 26S and was less susceptible to long-branch attraction. The weaker support of the 26S results compared with 16S may be due to a high rate of substitution determined lambda for the 26S sequence of these species. Therefore, 16S phylogenies were used as the basis for choosing the quartet.

Incongruent QuartOPs

The sequences with the highest bit score from each of the compared sequences together with the query sequence from *M. cuprina* were used to form a QuartOP. After alignment with ClustalW and bootstrap construction using RaxML a barycentric graph was constructed. The barycentric graph is useful for graphically showing phylogenic grouping. The graph showed strong support for the topological grouping of *M. cuprina* with *A. hospitalis* and *P. islandicum* with *V.moutnovskia.* This topology garnered support from 93 % of the QuartOPs and very strong support from 85 % (bootstrap branch support greater than 95). However, there were 27 QuartOPs that strongly supported other topologies (Fig. 2) which could indicate HGT. While any of these QuartOPs could be further analyzed for HGT, for the purposes of this investigation a QuartOP with strongest support was utilized for further analysis.

HGT of a Ferredoxin Gene

The QuartOP that was chosen strongly supported the topology of (*A. hospitalis, V. moutnovskia*) and (*M. cuprina, P. islandicum*) (B in Fig. 2). This topology was in opposition with the 93 % majority (A in Fig. 2), as well as 16S (Fig. 3). Assuming that the support for a conflicting topology represents a real case of genomic mosaicism, this could be due to either HGT or gene loss (Fig. 4). The two cases can be differentiated by comparing the topologies of the ferredoxin phylogeny. It would be expected that in the case of HGT, the organism that putatively received the transferred gene would cluster closely with the other group. The ferredoxin phylogeny includes all 35 taxa and the 16S phylogeny provides a good representation of the overall organism phylogeny. If gene loss in *M. cuprina* occurred, the *M. cuprina* ortholog in the QuartOP would be an unrecognized paralog of some distant gene. This would place *M. cuprina* on a long-branch, which could have led it to group with either *V. moutnovskia* or *P. islandicum* due to

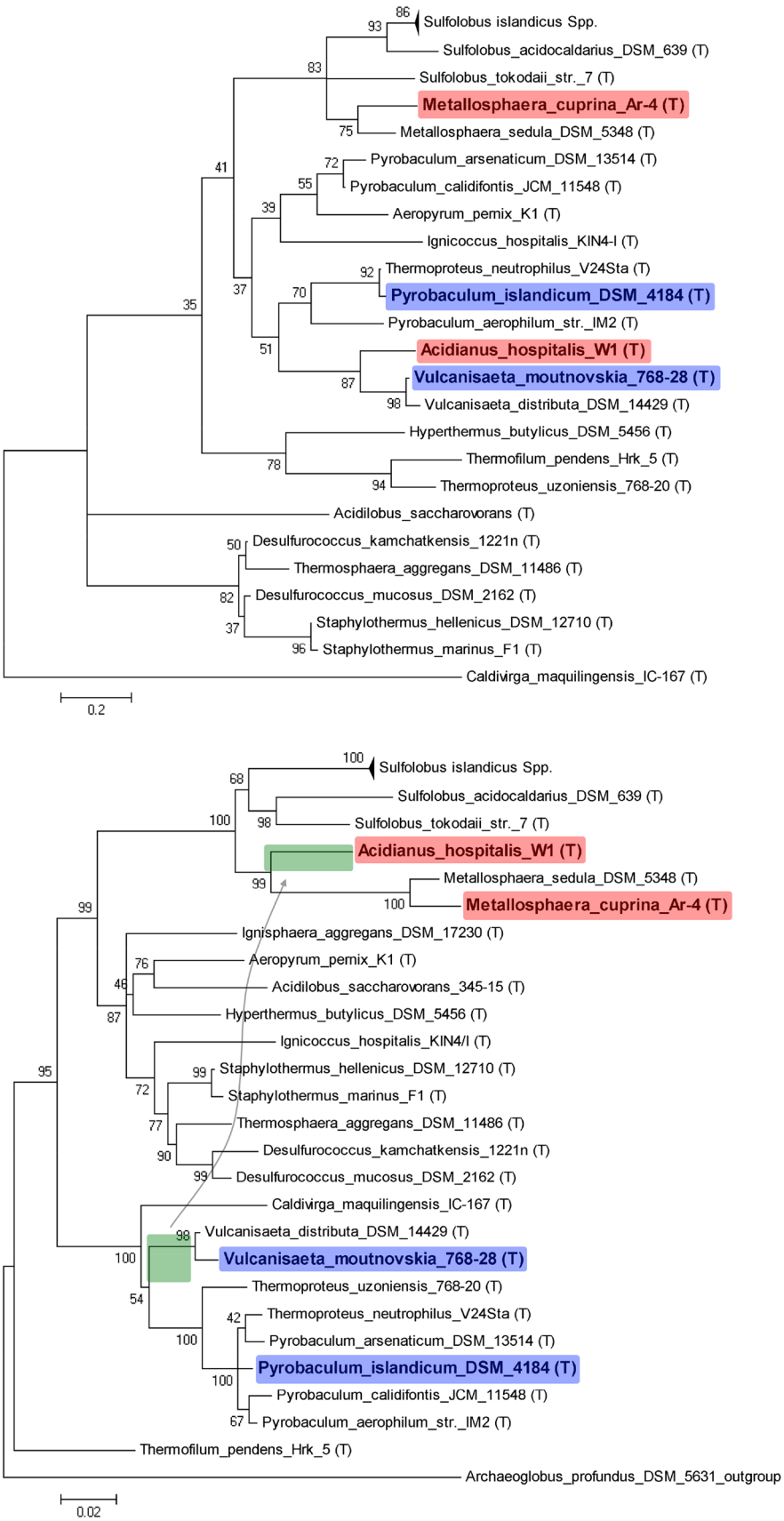

Fig. 3 Comparison of ferredoxin-related gene phylogeny (*upper tree*) with that of 16S (*lower tree*). *A. hospitalis* falls in with *V. moutnovskia*, suggesting HGT. The suspected HGT event is represented visually by a *trace line* on the 16S phylogeny

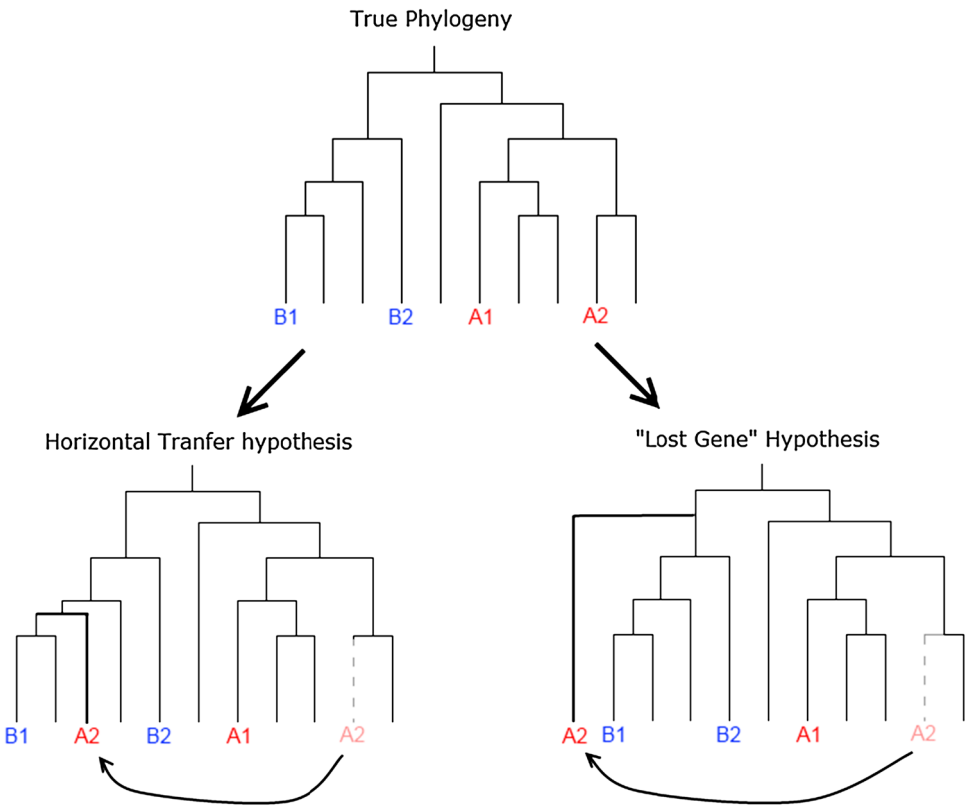

Fig. 4 Schematic concept of genomic mosaicism caused by HGT or gene loss. The transfer or loss of a suspected gene can lead to phylogenies (*bottom*) that differ from the "true" or assumed organism phylogeny (*top*). The phylogeny of orthologous genes with loss from A2 (at *low right*) would differ from that if instead A2 received a gene from B1 (*lower left*). HGT caused A2 grouping with B1 after splitting from B2

long-branch attraction. Alternatively, if the gene was transferred from *V. moutnovskia* to *A. hospitalis*, *A. hospitalis* would fall with *Vulcaniseata* after *V. moutnovskia* branched off from *P. islandicum*.

Comparing the ferredoxin gene phylogeny with the 16S phylogeny, *M. cuprina* maintains its relative place, while *A. hospitalis* groups with *V. moutnovskia* (Fig. 3). This comparison suggests that HGT is more likely than gene loss. *A. hospitalis* is not on a long-branch and no additional taxa are included between *V. moutnovskia* and *P. islandicum.* These results suggest HGT from *V. moutnovskia* to *A. hospitalis*. If gene loss were to have occured, we would have expected *A. hospitalis* to be separated from *V. moutnovskia* and *P. islandicum* with other species interleaved. Based on grouping, the relative chronological order of the transfer event can also be inferred. Since the ferredoxin in *A. hospitalis* grouped with *V. moutnovskia* and not *P. islandicum,* the gene was transferred after the *V. moutnovskia* lineage diverged from *P. islandicum*. Since the ferredoxin was transferred to *A. hospitalis* and neither of the two *Metallosphaera*, the ferredoxin gene was received after *A. hospitalis* diverged from the *Metallosphaera* lineage. MEGA was used to reconstruct the full 35 species phylogenies due to the stronger graphing support and to corroborate the findings of RAxML.

Discussion

HGT occurs more often in microorganisms that reside in similar environments (Zhaxybayeva and Gogarten 2003). Even between the Bacteria and Archaea domains, HGT can occur with high frequency (Nelson et al. 1999; Koonin et al. 2001). Nelson et al. (1999) found that the novel bacterial species *Thermotoga maritima* had approximately 24 % of genes more similar to Archaea than to its closest bacterial relatives.

Vulcanisaeta moutnovskia and *A. hospitalis* are acidophilic thermophiles and are found in hot springs around the world (You et al. 2011; Gumerov et al. 2011; Mavromatis et al. 2010). It is not unreasonable that a HGT event was observed between the *V. moutnovskia* and *A. hospitalis* lineages, as high rates of HGT has been reported in Crenarchaeota and many are extremophiles that reside in the same environment (DeLong 1992; Fuhrman 1992; Barns et al. 1996). However, there are still open questions, for example, which genes are more frequently horizontally transferred, and to what extent HGT takes place within the phylum or with other species outside the phylum.

From single gene sequences extracted from the genome annotation files, we have observed 27 homologous sequences that are putative HGTs (B and C in Fig. 2) and

the results have demonstrated the phylogenetic relationships among four taxa, given the support of the majority of QuartOPs. The use of bootstrap for the branch support values, rather than, for example, maximum likelihood posterior probability, leads to a conservative estimation of the extent of HGT (Zhaxybayeva and Gogarten 2003). Manual inspection of the alignment and incorporation of one specific QuartOP (a ferrodoxin related gene) into a larger phylogeny has confirmed that this gene was indeed horizontally transferred from the *V. moutnovskia* lineage to *A. hospitalis*.

The results have shown that quartet analysis can be used in screening homologous sequences for putative HGTs and is useful in visually describing genomic mosaicism and HGT within four taxa. However, each putative HGT must be incorporated into a larger phylogeny to differentiate HGT from unrecognized paralogy, composition bias, or phylogenetic reconstruction artifacts. A putative HGT may also be due to composition bias (a type of convergent evolution), providing a false positive, as pointed out by Jermiin et al. (2004).

Some proteins may be excluded due to lack of orthologs in all four species in a quartet. Because an ortholog must be found within all four species of a quartet to become a QuartOP, the analysis often loses half or more of a given taxon's genes. The analysis done by Zhaxybayeva and Gogarten (2002) demonstrated this problem, finding taxa quartets that had as few as 82 genes in QuartOPs. This is a potentially larger problem than a first order estimate based purely on percentages, since variable presence of a gene in various taxa is indicative and characteristic of HGT (Gogarten and Townsend 2005).

Compared with standard phylogenetic techniques, quartet analysis is better suited for finding putative HGT events. Creating standard phylogenies for comparative analysis using total genomes are computationally demanding. Zhaxybayeva et al. (2006) and Kubatko and Degnan (2007) discuss the common technique of concatenating sequences to help resolve phylogenetic discrepancies and the use of the resulting phylogenies as a reference to which to base the level of HGT. Kubatko and Degnan (2007) analyze the concatenation approach by simulation and show that it performs poorly in predicting and supporting true topologies. Zhaxybayeva et al. (2006) point out that concatenation requires the assumption that the concatenated genes have a uniform phylogenetic history. No such assumptions are required - when using quartet analysis.

Zhaxybayeva et al. (2004) discusses a related technique, the bipartition method, in which they used "QuintOps" instead of "QuartOps." Here, they took a similar approach of finding orthologs between genomes, by performing BLAST of ORFs against the genomes of the other species. They used this method to help determine tree topologies that were left unresolved by standard phylogenetic techniques, by looking at "plurality" (i.e., majority) of supported quintOps.

Quartet analysis can show relationships among only four species. It cannot say anything about the other 31 Crenarchaeota. Limited computational power potentially plays a large role in this type of analysis for large groups since the number of possible quartets grows exponentially as more species are added. It may simply not always be practical or even useful to comprehensively analyze large data sets. Furthermore, not all quartets are equal. A quartet where all species are very closely related may not provide any additional information of interest; there may not be enough divergence between the species in the quartet to resolve phylogenetic topology. Furthermore, if one species in a quartet is substituted for another that is a very close neighbor, this new quartet might not provide any additional information. It is noteworthy that substituting *V. moutnovskia* for *V. distributa* would have found the same ferredoxin HGT event (data not shown). Other quartets may give faulty information, if each of the taxa is on long-branches that separate them from the other three species.

Conclusions

Analysis by quartets has not yet been performed on any group in the phylum Crenarchaeota. The results have shown that quartet analysis can be used to screen homologous sequences for putative HGTs and is useful in visually describing genomic mosaicism and HGT within Crenarchaeota species and taxa in general. It should be possible to create a mechanical algorithm to choose informative quartets based on a global phylogenetic topology, such as one based on 16S, and use those quartets as comprehensive representatives of the genomic mosaicism of the whole group. Such an algorithm does not yet exist, but it seems inevitably possible.

Acknowledgments We thank Daniel Rubinoff for helpful discussions. The work was supported in part by the U.S. ONR HEET (N00014-11-1-0391), National Center for Research Resources (5P20RR016467-11) and National Institute of General Medical Sciences (8 P20 GM103466-11) from the National Institutes of Health.

References

Barns SM, Delwiche CF, Palmer JD, Pace NR (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci USA 93:9188–9193

DeLong EF (1992) Archaea in coastal marine environments. Proc Natl Acad Sci USA 89:5685–5689

Fuhrman JA (1992) Novel major archaebacterial group from marine plankton. Nature 356:148–149

Garcia-Vallvé S, Romeu A, Palau J (2000) Horizontal gene transfer in bacterial and archaeal complete genomes. Genome Res 10: 1719–1725

Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol 3:679–687

Gumerov VM, Mardanov AV, Beletsky AV et al (2011) Complete genome sequence of "*Vulcanisaeta moutnovskia*" strain 768-28, a novel member of the hyperthermophilic crenarchaeal genus *Vulcanisaeta*. J Bacteriol 193:2355–2356

Jermiin LS, Ho SY, Ababneh F et al (2004) The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst Biol 53:638–643

Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification 1. Ann Rev Microbiol 55:709–742

Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst Biol 56:17–24

Kunin V, Goldovsky L, Darzentas N, Ouzounis CA (2005) The net of life: reconstructing the microbial phylogenetic network. Genome Res 15:954–959

Kurland CG, Canback B, Berg OG (2003) Horizontal gene transfer: a critical view. Proc Natl Acad Sci USA 100:9658–9662

Liu L-J, You X-Y, Zheng H et al (2011) Complete genome sequence of *Metallosphaera cuprina*, a metal sulfide-oxidizing archaeon from a hot spring. J Bacteriol 193:3387–3388. doi:10.1128/JB.05038-11

Mavromatis K, Sikorski J, Pabst E et al (2010) Complete genome sequence of *Vulcanisaeta distributa* type strain (IC-017T). Stand Genomic Sci 3:117

Nelson KE, Clayton RA, Gill SR et al (1999) Evidence for lateral gene transfer between *Archaea* and bacteria from genome sequence of *Thermotoga maritima*. Nature 399:323–329

Ragan MA (2001) Detection of lateral gene transfer among microbial genomes. Curr Opin Genet Dev 11:620–626

Reno ML, Held NL, Fields CJ et al (2009) Biogeography of the *Sulfolobus islandicus* pan-genome. Proc Natl Acad Sci USA 106:8605–8610

Ribeiro S, Golding GB (1998) The mosaic nature of the eukaryotic nucleus. Mol Biol Evol 15:779–788

Salzberg SL, White O, Peterson J, Eisen JA (2001) Microbial genes in the human genome: lateral transfer or gene loss? Science 292:1903–1906

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690

Strimmer K, Von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci USA 94:6815–6819

Tamura K, Peterson D, Peterson N et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Wolf YI, Koonin EV (2012) A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes. Genome Biol Evol 4:1286–1294

You X-Y, Liu C, Wang S-Y et al (2011) Genomic analysis of *Acidianus hospitalis* W1 a host for studying crenarchaeal virus and plasmid life cycles. Extremophiles 15:487–497

Young JM (2001) Implications of alternative classifications and horizontal gene transfer for bacterial taxonomy. Int J Syst Evol Microbiol 51:945–953

Zhaxybayeva O, Gogarten JP (2002) Bootstrap, Bayesian probability and maximum likelihood mapping: exploring new tools for comparative genome analyses. BMC Genom 3:4

Zhaxybayeva O, Gogarten JP (2003) An improved probability mapping approach to assess genome mosaicism. BMC Genom 4:37

Zhaxybayeva O, Hamel L, Raymond J, Gogarten JP (2004) Visualization of the phylogenetic content of five genomes using dekapentagonal maps. Genome Biol 5:R20

Zhaxybayeva O, Gogarten JP, Charlebois RL et al (2006) Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. Genome Res 16:1099–1108