Gene 518 (2013) 483–488

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Letter to the Editor

Unrecognized fine-scale recombination can mimic the effects of adaptive radiation

Weilong Hao *

Department of Biological Sciences, Wayne State University, Detroit, MI 48202, USA

ARTICLE INFO

Article history:
Accepted 25 December 2012
Available online 18 January 2013

Keywords:
Adaptive evolution
Homologous recombination
Horizontal gene transfer
Incongruence
oprF
Pseudomonas

ABSTRACT

Gene sequences can undergo accelerated nucleotide changes and rapid diversification. The rapid sequence changes can then potentially lead to phylogenetic incongruence. Recently, Bodilis et al. (2011) observed artificial phylogenetic incongruence using the *Pseudomonas* surface protein gene *oprF*, and hypothesized that it was the result of a long-branch attraction artifact ultimately caused by adaptive radiation. In this study, an alternative hypothesis, namely fine-scale recombination, was tested on the same dataset. The results reveal that regions in *oprF* are of different evolutionary origins, and the mosaic gene structure resulted in confounding phylogenetic signals. These findings demonstrate that unrecognized fine-scale recombination can confound the phylogenetic interpretation and emphasize the limitation of using whole genes as the unit of phylogenetic analysis.

Recently, Bodilis et al. (2011) reported a phylogenetic incongruence in the position of *Pseudomonas fluorescens* based on an analysis using the outer membrane protein gene OprF. OprF functions as a porin, a root adhesion protein, that acts as an environmental sensor (Bodilis et al., 2006; Woodruff and Hancock, 1989). In the *oprF* phylogeny, the *P. fluorescens* group was split into two distinct groups, termed fluorescens 1 and 2 o-subclusters (or fluo1 and fluo2). The fluo2 subcluster grouped with two members (*P. mosselii* and *P. plecoglossicida*) of the *P. putida* group (abbreviated as Monofluo2putida in Bodilis et al., 2011, also shown in Supplementary Fig. S.1). The observed phylogenetic incongruence between the *oprF* phylogeny and the species tree based on four concatenated housekeeping genes was *not* statistically significant (Bodilis et al., 2011). A monophyletic *P. fluorescens* group (Monofluo), consistent with the species tree, was obtained when using partitioned analysis by codon position, or without partitioned analysis but excluding *P. mosselii* and *P. plecoglossicida*. Bodilis et al. (2011) argued that there is a link between the *oprF* phylogeny and the habitat of the bacterial strains, and remarked that the attraction between fluo2 and the *P. mosselii*–*P. plecoglossicida* clade was associated with adaptive radiation.

One alternative explanation of phylogenetic incongruence could be horizontal gene transfer (HGT), which is a major force in bacterial genome evolution (Beiko et al., 2005; Gogarten et al., 2002; Ochman et al., 2000). HGT can occur at the subgenic level (Chan et al., 2009a). To date, several surface proteins have been found to be subject to HGT at the subgenic level. For instance, the *Wolbachia* surface protein gene *wsp* contains four hypervariable regions, and extensive exchange of these four hypervariable regions occurred among *Wolbachia* strains infecting arthropods from different supergroups (Baldo et al., 2005). In the malaria parasite, recombination has been found to play an important role in the evolution of the merozoite surface antigen MSP-3alpha (Mascorro et al., 2005). In the human pathogen *Neisseria meningitidis*, recurrent intra-genic recombination has been found in at least three genes (*ctrC*, *ctrD*, and *ctrG*) within the capsule gene cluster and one vaccine target gene (NHBA) of the newly developed 4CMenB vaccine (Hao et al., 2011). These findings highlight the need to assess the existence of intra-genic recombination in the *oprF* sequences of *Pseudomonas*.

Recombination test analysis was performed on the 71 *oprF* sequences of *Pseudomonas* published by Bodilis et al. (2011). DNA sequences were first translated to protein sequences, aligned using MUSCLE (Edgar, 2004), and then converted back to the corresponding nucleotide sequences using Seaview (Gouy et al., 2010). At the ends of the full alignment, regions that were not present in all 71 species were trimmed. In Bodilis et al. (2011), analysis was conducted on the conserved sequence blocks filtered by Gblocks (Castresana, 2000). In this study, both the unfiltered original alignment and conserved blocks filtered by Gblocks were examined. In addition, another set of conserved blocks were selected using BMGE (Criscuolo and Gribaldo, 2010), as the Gblocks program has been suggested to be too conservative (Criscuolo and Gribaldo, 2010). Both the Gblocks and BMGE programs were applied on the translated protein alignment and corresponding DNA alignments were extracted using in-house PERL scripts. All sequence alignments generated in the study are available at https://sites.google.com/site/haowlab/. Phylogeny reconstruction was conducted using PhyML (Guindon et al., 2010) with a GTR + Γ + 1 substitution model, eight discrete categories for the Γ distribution, and 100 bootstrap iterations unless otherwise specified.

Abbreviations: AU test, approximately unbiased test; FDR, false discovery rate; HGT, horizontal gene transfer.

* Tel.: +1 313 577 6450.
E-mail address: haow@wayne.edu.

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.12.107

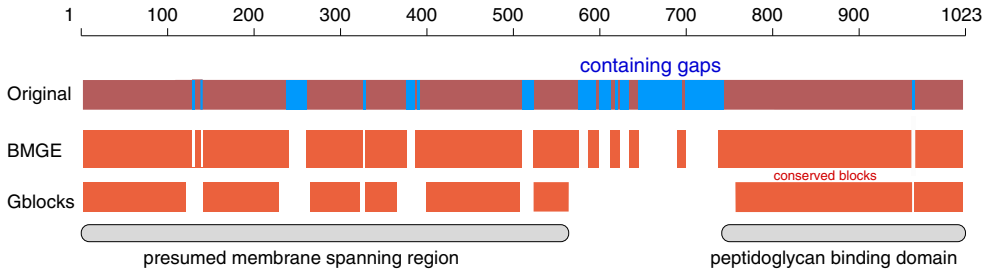

Fig 1. Distribution of gaps and conserved blocks in the sequence alignment. The 5′-end of the sequence (N-terminal of the corresponding protein sequence) matched the functional domain *pfam05736* in the presumed membrane spanning region, and the 3′-end of the sequence matched the peptidoglycan binding domain *pfam00691*. The gaps in the original sequence alignment were shown in blue, and the selected conserved blocks by BMGE and Gblocks were shown in red.

Consistent with Bodilis et al. (2011), the alignment filtered by Gblocks showed a Monofluo2putida relationship (Supplementary Fig. S.1). A Monofluo relationship was obtained (Supplementary Figs. S.2 and S.3), when the unfiltered full-length alignment and the BMGE-filtered alignment were analyzed. Based on the AU test (Shimodaira, 2002), Monofluo was not significantly different from Monofluo2putida using any of the three whole-sequence alignments. These results thus reveal that the attraction between the fluorescens 2 subcluster and the *P. mosselii*–*P. plecoglossicida* clade is weak.

There are two main functional domains in *oprF*, the presumed membrane spanning region and the peptidoglycan binding domain (Fig. 1). There are a number of insertions/deletions (indels) between the two domains, which were observed as gaps in the sequence alignment. It is clear that BMGE removed most nucleotide sites containing gaps. Gblocks further removed many nucleotide sites flanking the gaps, and as the result, the Gblocks-filtered alignment did not contain nucleotide sites from the interdomain region (Fig. 1).

In the next step, phylogenetic incongruence was investigated using 100-nucleotide sliding windows on both the BMGE-filtered sequence alignment (Fig. 2) and the Gblocks-filtered sequence alignment (Fig. 3). On each sliding window, the AU test (Shimodaira, 2002) was conducted to examine the difference between Monofluo and Monofluo2putida (shown in Supplementary Fig. S.3). Among the 77 sliding windows on the BMGE-filtered sequence alignment, 12 rejected Monofluo at $P<0.05$ and 14 rejected Monofluo2putida at $P<0.05$. These 26 windows are clustered in the four regions (highlighted in Fig. 2A). After correcting for multiple tests using the False Discovery Rate (FDR), 12 sliding windows in the *oprF* alignment remained significant with 11 of them rejecting Monofluo2putida. Furthermore, the number of sliding windows suggesting phylogenetic incongruence at $P<0.05$ was much larger than the number estimated from the permutated sequence alignment (Supplementary Fig. S.4). Similarly, among the 64 sliding windows on the Gblocks-filtered sequence alignment, 13 rejected Monofluo at $P<0.05$, 5 rejected Monofluo2putida at $P<0.05$. These windows are clustered in three regions (highlighted in Fig. 3A). The difference between Figs. 2A and 3A was mainly due to the lack of the interdomain region in the Gblocks-filtered sequence alignment (Fig. 1). It is important to note that the phylogenetic incongruence between the two alternative topologies (Monofluo vs. Monofluo2putida) was significant in many sliding windows but not in the whole-sequence alignment. It is possible that the combination of conflicting signals has weakened the phylogenetic signal. In fact, when the Gblocks-filtered sequence alignment was divided by two according to the functional domains (Fig. S.5), the first half of the alignment rejected Monofluo at $P=0.016$, and the second half of the alignment rejected Monofluo2putida at $P=0.001$. When each half of the sequence data was used to estimate trees independently, the supported trees were significantly incongruent ($P<0.001$, Supplementary Fig. S.5). These results suggest that the phylogenetic incongruence between Monofluo and Monofluo2putida is just a fraction of the total phylogenetic incongruence in *oprF*.

Further efforts were made to examine whether the observed Monofluo2putida relationship was actually due to the long branch attraction. The averaged branch length for fluo2 and *putida* was calculated for each sliding window based on pairwise distance between groups (Figs. 2B and 3B). In Fig. 2B, the longest branch for fluo2 was observed in the sliding window around nucleotide 520, and this window rejected Monofluo2putida ($P=0.012$). The longest branch length for *putida* was observed in the sliding window around nucleotide 310, but the sliding window did not significantly differentiate between Monofluo and Monofluo2putida. In Fig. 3B, neither the longest branch for fluo2 (around nucleotide 210) nor for putida (around nucleotide 260) showed significant difference between Monofluo and Monofluo2putida. Thus, it seems premature to conclude that the observed Monofluo2putida relationship based on the Gblocks filtered sequence alignment was solely due to the long branch attraction.

The Phi test (Bruen et al., 2006) was performed on each sliding window (Figs. 2C and 3C). Among the 77 sliding windows on the BMGE-filtered alignment, 59 windows had P-values <0.05 and 57 windows remained significant after correcting for multiple tests using the False Discovery Rate (FDR). Similarly, among the 64 sliding windows on the Gblocks filtered alignment, 41 windows had P-values <0.05 and 40 windows remained significant after correcting for multiple tests using FDR. Importantly, the second region that significantly rejected Monofluo consistently showed very low P-values for recombination. This suggests that some of the signal for phylogenetic incongruence in *oprF* was caused by recombination. It is worth noting that the Phi test examines recombination only within each sliding window and any gene transfer or gene conversion in a larger stretch beyond the boundaries of a sliding window was undetectable. Therefore, the extent of recombination in *orpF* was therefore underestimated. Furthermore, the recombinant sequence regions are not randomly distributed across the *Pseudomonas* phylogeny. Analysis with GENECONV (Sawyer, 1989) (Supplementary

Fig. 2. Association among phylogenetic incongruence, sequence divergence, and homologous recombination based on the sequence alignment filtered by BMGE. A), P-values for phylogenetic incongruence between Monofluo (the *Pseudomonas fluorescens* group is monophyletic as shown in Fig. S.2) and Monofluo2putida (the *P. fluorenscens* 2 subcluster is included in the *P. putida* group) measured using different regions of the *oprF* gene in the AU test. The AU test was conducted in sliding windows of 100 nucleotides by sliding 10 nucleotides at a time. Among the total 77 sliding windows, 12 sliding windows support Monofluo (at $P<0.05$) and 12 sliding windows support Monofluo2putida (at $P<0.05$). Regions of these significant sliding windows were highlighted in horizontal bars in all panels. B), Averaged branch lengths of the fluo2 subcluster and the *P. putida* group. Each pairwise distance between groups is the sum of the two branches [e.g., fluo1fluo2 = b(fluo1) + b(fluo2)]. The branch length of each group could then be calculated from the averaged pairwise distance between groups. That is b(fluo2) = (fluo1fluo2 + fluo2putida − fluo1putida)/2, and b(putida) = (fluo1putida + fluo2putida − fluo1fluo2)/2. C), P-values for recombination measured by the PHI package. 57 of the 77 sliding windows have significant P-values for recombination after correcting for multiple tests using the False Discovery Rate (FDR). For presentation purpose, P-values when $<10^{-6}$ were shown as 10^{-6}. D), P-values for phylogenetic incongruence of the best tree obtained from each sliding window against either the Monofluo or Monofluo2putida topology. P-values when $<10^{-8}$ were shown as 10^{-8}.