Evolution and nomenclature of the trimethoprim resistant dihydrofolate (dfr) reductases

Mateusz Faltyn, Brian Alcock, and Andrew G. McArthur*

M.G. DeGroote Institute for Infectious Disease Research, Department of Biochemistry and Biomedical Sciences, DeGroote School of Medicine, McMaster University, Hamilton, Ontario L8S 4K1, Canada

*To whom correspondence should be addressed.

Contact: mcarthua@mcmaster.ca

Abstract

With the increasing use of genome sequencing as a surveillance tool for molecular epidemiology of antimicrobial resistance (AMR), databases and clear nomenclature for AMR gene families are critical. Due to the convoluted nomenclatural history of the integron-associated trimethoprim-resistant dihydrofolatereductase (dfr) gene family, we decided to conduct a literature review, comparative sequence analysis, and phylogenetic investigation of the dfr family, the results of which are presented here and available at the Comprehensive Antibiotic Resistance Database (CARD). Overall, literature review and phylogenetic analysis resolved gene name synonyms based on sequence. We recommend adoption of phylogenetic methods to help guide AMR gene naming efforts and relegation of misleading names to synonyms.

1 Introduction

With the increasing use of genome sequencing as a surveillance tool for molecular epidemiology of antimicrobial resistance (AMR), databases and clear nomenclature for AMR gene families are critical. Due to the convoluted nomenclatural history of the integron-associated trimethoprim-resistant dihydrofolate reductase (dfr) gene family, we decided to conduct a literature review, comparative sequence analysis, and phylogenetic investigation of the dfr family, the results of which are presented here and available at the Comprehensive Antibiotic Resistance Database (CARD). Overall, literature review and phylogenetic analysis resolved gene name synonyms based on sequence. We recommend adoption of phylogenetic methods to help guide AMR gene naming efforts and relegation of misleading names to synonyms.

2 Analysis

After extensive literature review and comparative sequence analysis (BLAST and multiple sequence alignment), we were able to identify 45 TMP resistance protein sequences using 55 different names (Table 1). Dihydrofolate reductase nomenclature for the most part is divided into two main branches (dfrAs and dfrBs), yet two protein sequences had histories of being named to either branch. As such, we performed a phylogenetic analysis of the dfr protein sequences, using ClustalW (13) to generate a multiple sequence alignment and RAxML (14) for protein phylogenetics under the JTT substitution model, a gamma distribution for among-site rate variation, proportion of invariable sites, and empirical amino acid frequencies. The resulting phylogenetic tree strongly supported separation of the dfrAs and dfrBs (Figure 1) and provided clear support for placement of dfrH / dfr2a / dfrB1, dfrA2d / dfrB4, and dfr-1c / dfrB5 within the dfrBs. We have thus relegated the names dfrH, dfr2A, dfrA2d, and dfr-1c to synonyms (Table 1). In addition, we discovered the names dfrB7 and dfrB8 described the same amino acid sequence and we relegated dfrB8 to a synonym (Table 1). Other names such as dfrA4, dfrA11, dfrA33, dfrH, and dfrJ could not be found by literature scan, while the paper describing dfrA2 had previously been retracted.

Table 1 provides CARD’s Antibiotic Resistance Ontology accession for each gene as well as the GenBank accession for each encoded protein sequence. Citations for the first report for each gene can be found in CARD, but CARD’s Resistomes & Variants data set (v3.0.2), which screens thousands of genome, plasmid, and whole genome shotgun assemblies for the presence of AMR genes, illustrates clearly the expansion of TMP resistant dihydrofolate reductase to a broad range of pathogens, most notably for dfrA8, dfrA10, and dfrA14 (https://card.mcmaster.ca/prevalence). Altogether between the published literature and CARD’s surveillance efforts, TMP resistant dihydrofolate reductases are found at least 35 pathogens.

3 Conclusions
As the rate of genome and metagenome sequencing accelerates, new AMR genes and variants are going to be found more frequently, presenting a challenge for naming of genes. Yet clear nomenclature will be critical for molecular epidemiological efforts and data harmonization among different agencies. We recommend adoption of phylogenetic methods to help guide naming efforts and relegation of misleading names to synonyms. We note from our phylogenetic analysis that the dfrC – dfrK genes are more closely related to dfrAs than dfrBs and do not form novel lineages. This suggests the dihydrofolate reductases of the gram-positive pathogens Listeria, Staphylococcus, Streptococcus, and Enterococcus (Table 1) share a common evolutionary history with the dfrAs (found in gram-negative pathogens), likely through rare horizontal gene transfer events (Figure 1). As such, examinations of sequence similarity, nomenclature, and evolutionary relationships are necessary to provide a common interpretative framework and can lead to important epidemiological implications.

4 Acknowledgements

We thank Dr. Daniel Haft (National Center for Biotechnology Information, National Institutes of Health, USA) for assistance with dfrA18 and dfrA19 curation (github.com/arpcard/amr_curation/issues/1). This research was funded by the Canadian Institutes of Health Research (PJ1-156214) to A.G.M., who also holds a Cisco Research Chair in Bioinformatics, supported by Cisco Systems Canada, Inc. Computer resources were supplied by the McMaster Service Lab and Repository computing cluster, funded in part by grants to A.G.M. from the Canadian Foundation for Innovation (34531).

References


Figure 1. Phylogenetic analysis of TMP resistant dihydrofolate reductases, with branch lengths representative of evolutionary distance and nodes labelled with bootstrap support.
Table 1. Biocuration of TMP resistant dihydrofolate reductase protein sequences.

<table>
<thead>
<tr>
<th>Dihydrofolate Reductase</th>
<th>ARO Accession</th>
<th>GenBank Accession</th>
<th>Synonyms</th>
<th>First Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>dfrA1</td>
<td>ARO:3002854</td>
<td>CAC19929.1</td>
<td>dfrA1</td>
<td>Escherichia coli integron</td>
</tr>
<tr>
<td>dfrA3</td>
<td>ARO:3003105</td>
<td>AAA25550.1</td>
<td>dfr6</td>
<td>Vibrio cholerae integron</td>
</tr>
<tr>
<td>dfrA5</td>
<td>ARO:3002861</td>
<td>ABB89122.1</td>
<td>Vibrio cholerae integron</td>
<td></td>
</tr>
<tr>
<td>dfrA6</td>
<td>ARO:3004547</td>
<td>ACY06082.1</td>
<td>dfr6</td>
<td>Vibrio cholerae integron</td>
</tr>
<tr>
<td>dfrA7</td>
<td>ARO:3002862</td>
<td>ACS44716.1</td>
<td>Escherichia coli integron</td>
<td></td>
</tr>
<tr>
<td>dfrA8</td>
<td>ARO:3002863</td>
<td>AHY80711.1</td>
<td>Salmonella enterica transposon</td>
<td></td>
</tr>
<tr>
<td>dfrA9</td>
<td>ARO:3004548</td>
<td>CAA40897.1</td>
<td>dfr9, dfrIX</td>
<td>Escherichia coli plasmid</td>
</tr>
<tr>
<td>dfrA10</td>
<td>ARO:3003011</td>
<td>AHG97174.1</td>
<td>Klebsiella pneumoniae integron</td>
<td></td>
</tr>
<tr>
<td>dfrA12</td>
<td>ARO:3002858</td>
<td>AHW42429.1</td>
<td>Vibrio cholerae integron</td>
<td></td>
</tr>
<tr>
<td>dfrA13</td>
<td>ARO:3003012</td>
<td>CCA090683.1</td>
<td>Escherichia coli integron</td>
<td></td>
</tr>
<tr>
<td>dfrA14</td>
<td>ARO:3002859</td>
<td>ACE32277.1</td>
<td>Escherichia coli integron</td>
<td></td>
</tr>
<tr>
<td>dfrA15</td>
<td>ARO:3003013</td>
<td>AHB39758.1</td>
<td>Vibrio cholerae integron</td>
<td></td>
</tr>
<tr>
<td>dfrA16</td>
<td>ARO:3003014</td>
<td>AAK60186.1</td>
<td>Salmonella enterica integron</td>
<td></td>
</tr>
<tr>
<td>dfrA17</td>
<td>ARO:3002860</td>
<td>ABQ91835.1</td>
<td>Escherichia coli integron</td>
<td></td>
</tr>
<tr>
<td>dfrA18</td>
<td>ARO:3004568</td>
<td>AAK64581.1</td>
<td>dfr18</td>
<td>Vibrio cholerae integron</td>
</tr>
<tr>
<td>dfrA19</td>
<td>ARO:3003015</td>
<td>CAC81324.1</td>
<td>Klebsiella pneumoniae integron</td>
<td></td>
</tr>
<tr>
<td>dfrA20</td>
<td>ARO:3003016</td>
<td>CAE53424.1</td>
<td>Pasteurella multocida plasmid</td>
<td></td>
</tr>
<tr>
<td>dfrA21</td>
<td>ARO:3003017</td>
<td>CAP96959.1</td>
<td>Salmonella enterica integron</td>
<td></td>
</tr>
<tr>
<td>dfrA22</td>
<td>ARO:3003018</td>
<td>CAX16467.1</td>
<td>Salmonella enterica integron</td>
<td></td>
</tr>
<tr>
<td>dfrA23</td>
<td>ARO:3003019</td>
<td>CAG4233.2</td>
<td>Salmonella enterica integron</td>
<td></td>
</tr>
<tr>
<td>dfrA24</td>
<td>ARO:3002856</td>
<td>CAP9385.1</td>
<td>Escherichia coli integron</td>
<td></td>
</tr>
<tr>
<td>dfrA25</td>
<td>ARO:3003020</td>
<td>ABB71176.1</td>
<td>Salmonella enterica integron</td>
<td></td>
</tr>
<tr>
<td>dfrA26</td>
<td>ARO:3002857</td>
<td>CAL48457.1</td>
<td>Escherichia coli integron</td>
<td></td>
</tr>
<tr>
<td>dfrA27</td>
<td>ARO:3004550</td>
<td>ACD45689.1</td>
<td>Vibrio cholerae integron</td>
<td></td>
</tr>
<tr>
<td>dfrA28</td>
<td>ARO:3004551</td>
<td>CAT00035.1</td>
<td>Aeromonas salmonicida integron</td>
<td></td>
</tr>
<tr>
<td>dfrA29</td>
<td>ARO:3004554</td>
<td>ANN25980.1</td>
<td>Escherichia coli integron</td>
<td></td>
</tr>
<tr>
<td>dfrA30</td>
<td>ARO:3004552</td>
<td>CAQ53499.2</td>
<td>Klebsiella pneumoniae integron</td>
<td></td>
</tr>
<tr>
<td>dfrA31</td>
<td>pending</td>
<td>BAD88719.1</td>
<td>Vibrio cholerae plasmid</td>
<td></td>
</tr>
<tr>
<td>dfrA32</td>
<td>ARO:3004555</td>
<td>ACZ52983.1</td>
<td>Salmonella enterica integron</td>
<td></td>
</tr>
<tr>
<td>dfrA34</td>
<td>pending</td>
<td>KTN70288.1</td>
<td>Salmonella enterica</td>
<td></td>
</tr>
<tr>
<td>dfrB1</td>
<td>ARO:3002864</td>
<td>AAY33960.1</td>
<td>dfr4, dfr2a</td>
<td>Bordetella bronchiseptica plasmid</td>
</tr>
<tr>
<td>dfrB2</td>
<td>ARO:3003021</td>
<td>FAA00064.1</td>
<td>Uncultured bacteria integron</td>
<td></td>
</tr>
<tr>
<td>dfrB3</td>
<td>ARO:3003022</td>
<td>ACR57831.1</td>
<td>Klebsiella oxtocu integron</td>
<td></td>
</tr>
<tr>
<td>dfrB4</td>
<td>ARO:3004498</td>
<td>AKF12264.1</td>
<td>dfrA2d</td>
<td>Escherichia coli integron</td>
</tr>
<tr>
<td>dfrB5</td>
<td>ARO:3004549</td>
<td>AAX46054.1</td>
<td>Pseudomonas aeruginosa integron</td>
<td></td>
</tr>
<tr>
<td>dfrB6</td>
<td>ARO:3003023</td>
<td>ADD00942.1</td>
<td>Salmonella enterica integron</td>
<td></td>
</tr>
<tr>
<td>dfrB7</td>
<td>ARO:3004556</td>
<td>ADB54781.1</td>
<td>Aeromonas hydrophila integron</td>
<td></td>
</tr>
<tr>
<td>dfrB9</td>
<td>pending</td>
<td>AML24034.1</td>
<td>Enterobacter cloacae integron</td>
<td></td>
</tr>
<tr>
<td>dfrC</td>
<td>ARO:3002865</td>
<td>AAO04716.1</td>
<td>Staphylococcus aureus chromosome</td>
<td></td>
</tr>
<tr>
<td>dfrD</td>
<td>ARO:3002866</td>
<td>AAA85213.1</td>
<td>Listeria monocytogenes plasmid</td>
<td></td>
</tr>
<tr>
<td>dfrE</td>
<td>ARO:3002875</td>
<td>EOD99669.1</td>
<td>Enterococcus faecalis chromosome</td>
<td></td>
</tr>
<tr>
<td>dfrF</td>
<td>ARO:3002867</td>
<td>AAD18086.1</td>
<td>Streptococcus pyogenes chromosome</td>
<td></td>
</tr>
<tr>
<td>dfrG</td>
<td>ARO:3002868</td>
<td>BAE15963.1</td>
<td>Staphylococcus aureus plasmid</td>
<td></td>
</tr>
<tr>
<td>dfrI</td>
<td>pending</td>
<td>ABQ40886.1</td>
<td>Versinia ruckert plasmid</td>
<td></td>
</tr>
<tr>
<td>dfrK</td>
<td>ARO:3002869</td>
<td>CBL04355.1</td>
<td>Staphylococcus aureus plasmid</td>
<td></td>
</tr>
</tbody>
</table>