 doi:10.20944/preprints201905.0137.v1

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) (1, 2), as well as the targeting of specific AMR genes by novel adjuvants (3), databases and clear nomenclature for AMR gene families is critical. The use of the same name for different sequences or different names for the same sequence hampers progress and it is the job of biocurators to set things right, much like their taxonomist progenitors. Toulouse et al. (4) recently characterized the activity of the integron-associated dihydrofolate reductase dfrB4 that confers resistance to trimethoprim and had to navigate it's convoluted nomenclatural history. In the process of curating this gene into the Comprehensive Antibiotic Resistance Database (5), we found it was not the only dfr sequence lacking clarity. We thus undertook a literature review, comparative sequence analysis, and phylogenetic investigation to better understand the evolutionary history and nomenclature of this AMR gene family.

The diaminopyrimidine trimethoprim (TMP) is a antimicrobial agent that was introduced in 1962 to treat infections in humans (6). TMP inhibits the activity of dihydrofolate reductase enzymes of eukaryotic and microbial cells (7) and when combined with sulfonamides (SULs) the two have a synergistic effect where TMP acts a SUL potentiator (8). The increased effectiveness and low cost of using TMPs and SULs led to their extensive application as drugs in the clinical treatment of infection (8, 9). Yet pathogens have evolved resistance to TMP via acquisition of alternate dihydrofolate reductases, often encoded on mobile genetic elements (i.e. antibiotic target replacement), particularly for *Escherichia coli* and *Klebsiella pneumoniae* (6, 9-11). This increase in the prevalence of trimethoprim resistant dihydrofolate reductase genes has resulted in a greater workload for primary healthcare and treatment failure in patients (12). In order to further understand TMP and SUL resistance and in turn reduce resistance levels, it is imperative to discover and correctly identify the various integron-encoded dihydrofolate reductases present in clinical pathogens and beyond.

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 dfrII / dfr2a / dfrB1, dfrA2d / dfrB4, and dfr-lie / dfrB5 within the dfrBs. We have thus relegated the names dfrII, dfr2A, dfrA2d, and dfr-lie 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 (PJT-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

1. **McArthur AG**, **Tsang KK**. 2017. Antimicrobial resistance surveillance in the genomic age. *Ann N Y Acad Sci* **1388**:78–91.

2. **McArthur AG**, **Wright GD**. 2015. Bioinformatics of antimicrobial resistance in the age of molecular epidemiology. *Curr Opin Microbiol* **27**:45–50.

3. **King AM**, **Reid-Yu SA**, **Wang W**, **King DT**, **de Pascale G**, **Strynadka NC**, **Walsh TR**, **Coombes BK**, **Wright GD**. 2014. Aspergillomarasmine A overcomes metallo-β-lactamase antibiotic resistance. *Nature* **510**:503–506.

4. **Toulouse JL**, **Edens TJ**, **Alejaldre L**, **Manges AR**, **Pelletier JN**. 2017. Integron-associated dfrB4, a previously uncharacterized member of the trimethoprim-resistant dihydrofolate reductase B family, is a clinically identified emergent source of antibiotic besistance. *Antimicrob Agents Chemother* **61**:1851–5.

5. **Jia B**, **Raphenya AR**, **Alcock B**, **Waglechner N**, **Guo P**, **Tsang KK**, **Lago BA**, **Dave BM**, **Pereira S**, **Sharma AN**, **Doshi S**, **Courtot M**, **Lo R**, **Williams LE**, **Frye JG**, **Elsayegh T**, **Sardar D**, **Westman EL**, **Pawlowski AC**, **Johnson TA**, **Brinkman FSL**, **Wright GD**, **McArthur AG**. 2017. CARD 2017: expansion and model-centric curation of the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* **45**:D566–D573.

6. **Huovinen P**, **Sundström L**, **Swedberg G**, **Sköld O**. 1995. Trimethoprim and sulfonamide resistance. *Antimicrob Agents Chemother* **39**:279–289.

7. **Burchall JJ**. 1979. The development of the diaminopyrimidines. *J Antimicrob Chemother* **5**:3–14.

8. **Bushby SR**, **Hitchings GH**. 1968. Trimethoprim, a sulphonamide potentiator. *Br J Pharmacol Chemother* **33**:72–90.

9. **Ho JM-W**, **Juurlink DN**. 2011. Considerations when prescribing trimethoprim-sulfamethoxazole. *Canadian Medical Association Journal* **183**:1851–1858.

10. **Brolund A**, **Sundqvist M**, **Kahlmeter G**, **Grape M**. 2010. Molecular characterisation of trimethoprim resistance in *Escherichia coli* and *Klebsiella pneumoniae* during a two year intervention on trimethoprim use. *PLoS ONE* **5**:e9233.

11. **Sköld O**. 2001. Resistance to trimethoprim and sulfonamides. *Vet Res* **32**:261–273.

12. **Butler CC**, **Hillier S**, **Roberts Z**, **Dunstan F**, **Howard A**, **Palmer S**. 2006. Antibiotic-resistant infections in primary care are symptomatic for longer and increase workload: outcomes for patients with *E. coli* UTIs. *Br J Gen Pract* **56**:686–692.

13. **Larkin MA**, **Blackshields G**, **Brown NP**, **Chenna R**, **McGettigan PA**, **McWilliam H**, **Valentin F**, **Wallace IM**, **Wilm A**, **Lopez R**, **Thompson JD**, **Gibson TJ**, **Higgins DG**. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948.

14. **Stamatakis A**. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2019 doi:10.20944/preprints201905.0137.v1

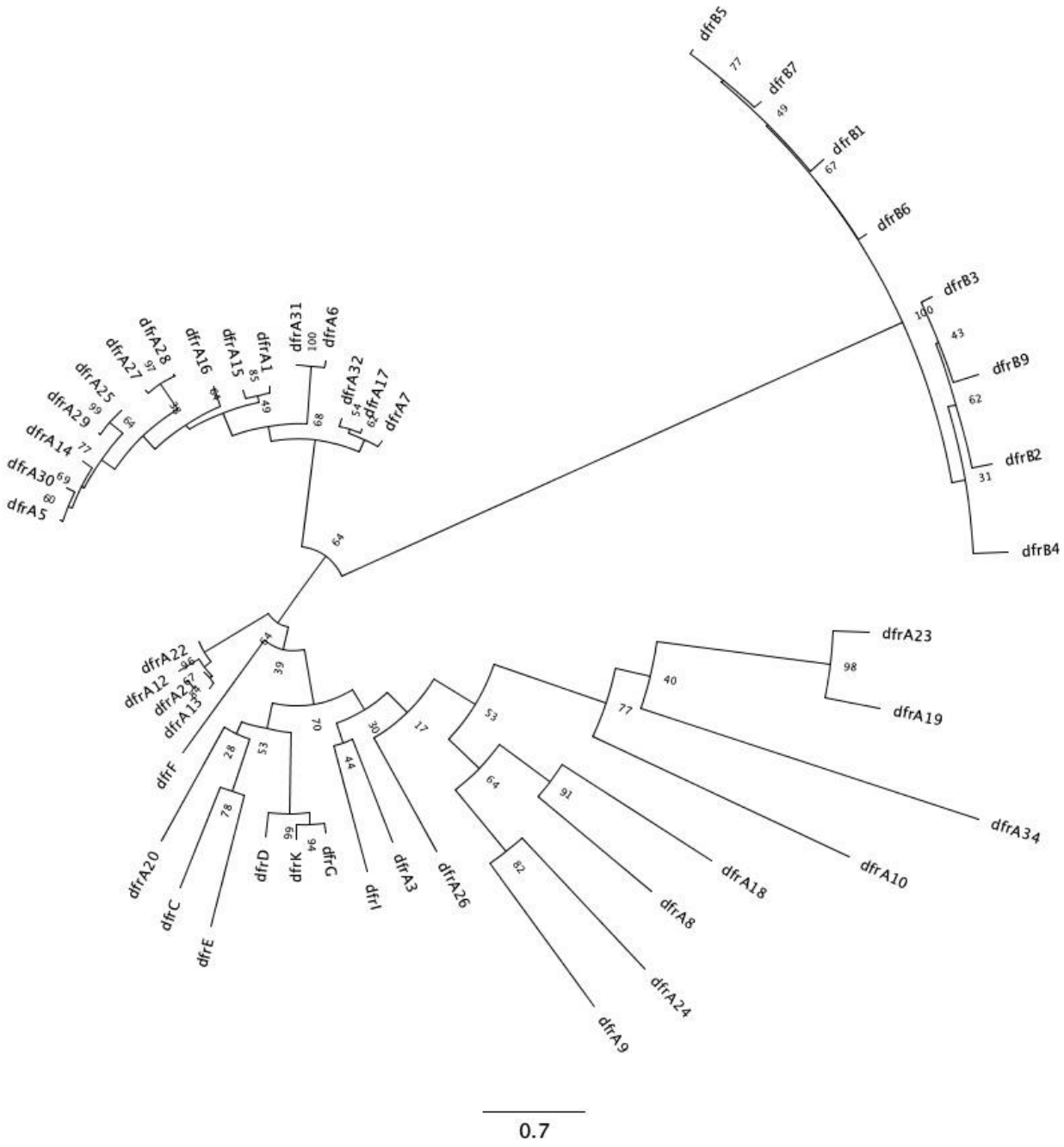

Figure 1. Phylogenetic analysis of TMP resistant dihydrofolate reductases, with branch lengths representative of evolutionary distance and nodes labelled with bootstrap support.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2019 doi:10.20944/preprints201905.0137.v1

Table 1. Biocuration of TMP resistant dihydrofolate reductase protein sequences.

Dihydrofolate Reductase	ARO Accession	GenBank Accession	Synonyms	First Reported
dfrA1	ARO:3002854	CAC19929.1	*dfr1*	*Escherichia coli* integron
dfrA3	ARO:3003105	AAA25550.1		*Escherichia coli* integron
dfrA5	ARO:3002861	ABB89122.1		*Vibrio cholerae* integron
dfrA6	ARO:3004547	ACY06082.1	*dfr6*	*Vibrio cholerae* integron
dfrA7	ARO:3002862	ACS44716.1		*Escherichia coli* integron
dfrA8	ARO:3002863	AHV80711.1		*Salmonella enterica* transposon
dfrA9	ARO:3004548	CAA40897.1	*dfr9, dfrIX*	*Escherichia coli* plasmid
dfrA10	ARO:3003011	AHG97174.1		*Klebsiella pneumoniae* integron
dfrA12	ARO:3002858	AHW42429.1		*Vibrio cholerae* integron
dfrA13	ARO:3003012	CAA90683.1		*Escherichia coli* integron
dfrA14	ARO:3002859	ACI32877.1		*Escherichia coli* integron
dfrA15	ARO:3003013	AHB39758.1		*Vibrio cholerae* integron
dfrA16	ARO:3003014	AAK60186.1		*Salmonella enterica* integron
dfrA17	ARO:3002860	ABG91835.1		*Escherichia coli* integron
dfrA18	ARO:3004568	AAK64581.1	*dfr18*	*Vibrio cholerae* integron
dfrA19	ARO:3003015	CAC81324.1		*Klebsiella pneumoniae* integron
dfrA20	ARO:3003016	CAE53424.1		*Pasteurella multocida* plasmid
dfrA21	ARO:3003017	CAP69659.1		*Salmonella enterica* integron
dfrA22	ARO:3003018	CAX16467.1		*Salmonella enterica* integron
dfrA23	ARO:3003019	CAG34233.2		*Salmonella enterica* integron
dfrA24	ARO:3002856	CAI99385.1		*Escherichia coli* integron
dfrA25	ARO:3003020	ABB71176.1		*Salmonella agona* integron
dfrA26	ARO:3002857	CAL48457.1		*Escherichia coli* integron
dfrA27	ARO:3004550	ACD45689.1		*Vibrio cholerae* integron
dfrA28	ARO:3004551	CAT00035.1		*Aeromonas salmonicida* integron
dfrA29	ARO:3004554	ANN23980.1		*Escherichia coli* integron
dfrA30	ARO:3004552	CAQ53849.2		*Klebsiella pneumoniae* integron
dfrA31	pending	BAD88719.1		*Vibrio cholerae* plasmid
dfrA32	ARO:3004555	ACZ52983.1		*Salmonella enterica* integron
dfrA34	pending	KTN70288.1		*Salmonella enterica*
dfrB1	ARO:3002864	AAY33960.1	*dfrII, dfr2a*	*Bordetella bronchispetica* plasmid
dfrB2	ARO:3003021	FAA00064.1		*Uncultured bacteria* integron
dfrB3	ARO:3003022	ACR57831.1		*Klebsiella oxytoca* integron
dfrB4	ARO:3004498	AKF12264.1	*dfrA2d*	*Escherichia coli* integron
dfrB5	ARO:3004549	AAX46054.1	*dfr-lie*	*Pseudomonas aeruginosa* integron
dfrB6	ARO:3003023	ADO00942.1		*Salmonella enterica* integron
dfrB7	ARO:3004556	ADB54781.1	*dfrB8*	*Aeromonas hydrophila* integron
dfrB9	pending	AGM20434.1		*Enterobacter cloacae* integron
dfrC	ARO:3002865	AAO04716.1		*Staphylococcus aureus* chromosome
dfrD	ARO:3002866	AAA85213.1		*Listeria monocytogenes* plasmid
dfrE	ARO:3002875	EOD99669.1		*Enterococcus faecalis* chromosome
dfrF	ARO:3002867	AAD01868.1		*Streptococcus pyogenes* chromosome
dfrG	ARO:3002868	BAE15963.1		*Staphylococcus aureus* plasmid
dfrI	pending	ABO40886.1		*Yersinia ruckeri* plasmid
dfrK	ARO:3002869	CBL80435.1		*Staphylococcus aureus* plasmid