


# Database of glutamate-gated chloride (GluCl) subunits across 125 nematode species: patterns of gene accretion and sequence diversification

Damien M. O'Halloran  \*

Department of Biological Sciences, The George Washington University, Washington, DC 20052, USA

\*Corresponding author: Department of Biological Sciences, The George Washington University, SEH 6000, 800 22nd Street NW, Washington, DC 20052, USA.  
Email: damienoh@gwu.edu

## Abstract

Glutamate-gated chloride channels belong to the Cys-loop receptor superfamily. Glutamate-gated chloride channels are activated by glutamate and form substrates for the antiparasitic drugs from the avermectin family. Glutamate-gated chloride channels are pentameric, and each subunit contains an N-terminal extracellular domain that binds glutamate and 4 helical transmembrane domains, which contain binding sites for avermectin drugs. In order to provide more insight into phylum-wide patterns of glutamate-gated chloride subunit gene expansion and sequence diversity across nematodes, we have developed a database of predicted glutamate-gated chloride subunit genes from 125 nematode species. Our analysis into this dataset described assorted patterns of species-specific glutamate-gated chloride gene counts across different nematodes as well as sequence diversity in key residues thought to be involved in avermectin binding.

**Keywords:** nematode; glutamate-gated chloride channel; GluCl; database; parasite; anthelmintic; ivermectin

## Introduction

Glutamate-gated chloride (GluCl) channels are pentameric ionotropic receptors belonging to the Cys-loop receptor superfamily. Each of the 5 subunits contain an N-terminal extracellular domain that binds Glutamate (Glu) and 4 helical transmembrane domains (TMs). The second TM domain (TM2) from each subunit lines a single channel pore, which opens in response to Glu binding (Hibbs and Gouaux 2011; Althoff et al. 2014). In the free-living nematode *Caenorhabditis elegans* 6 GluCl channel subunit genes have been identified: *glc-1* to *glc-4* and *avr-14* and *avr-15*; while the ruminant parasitic nematode *Haemonchus contortus* has 5 GluCl channel subunit genes: *Hco-glc-2*, *Hco-glc-3*, *Hco-avr-14*, *Hco-glc-5*, and *Hco-glc-6*. GluCl channels are also the substrate for the antiparasitic drug, ivermectin (IVM). IVM is hydrophobic and can reach the GluCl channel TMs to potentiate activity by binding within TM1 and TM2 of adjacent GluCl channel subunits while interacting with residues in the loop between TM2 and TM3 (Cully et al. 1994; Arena et al. 1995; Hibbs and Gouaux 2011; Calimet et al. 2013; Althoff et al. 2014). IVM binding stabilizes the TMs in an open configuration and increases the pore size, thereby increasing channel permeability (Althoff et al. 2014). The genes encoding distinct subunits are modulated differentially by both Glu and IVM: for example, the AVR-15 subunit of *C. elegans* (Dent et al. 1997), and the GluCl $\alpha$ 2B subunit of *H. contortus* (McCavera et al. 2009) can form homomeric channels that are sensitive to both IVM and Glu, while the *C. elegans* subunit GLC-1 forms homomeric receptors that can be activated by IVM but not by

Glu, and conversely homomeric GLC-2 channels can be activated by Glu but not by IVM (Cully et al. 1994; Vassilatis et al. 1997; Daeffler et al. 2014). Interestingly, heteromeric channels comprising GLC-1 and GLC-2 can be activated by both Glu and IVM independently (Cully et al. 1994; Vassilatis et al. 1997; Daeffler et al. 2014). In addition to subunit stoichiometry modulating IVM effects, the sensitivity of GluCl channels to IVM is also modulated by differences in sequence. A naturally occurring 4-amino acid deletion within the ligand-binding domain of GLC-1 in *C. elegans*, has been shown to confer resistance to abamectin, an avermectin class of anthelmintic (Ghosh et al. 2012). IVM resistance was first reported in *H. contortus* in 1979 (van Wyk and Malan 1988), and key residues have been identified within TM3 that are required for IVM sensitivity (Lynagh and Lynch 2010). These findings in *H. contortus* highlight the selection potential placed on GluCl subunit genes in response to IVM treatment, and more GluCl gene sequence data are required to understand population and phylum-wide patterns of sequence diversity. To get at this question of phylum-wide GluCl sequence diversity, we have developed a database of predicted GluCl subunit genes from 125 nematode species. All of these data are freely available for users to search and explore at the following URL: [http://ohalloranlab.net/nematode\\_glucl](http://ohalloranlab.net/nematode_glucl). Our analysis into this dataset described heterogeneous patterns of species-specific GluCl gene counts across different nematodes as well as sequence diversity in putative substrate residues for IVM binding.

**Received:** October 22, 2021. **Accepted:** December 14, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Materials and methods

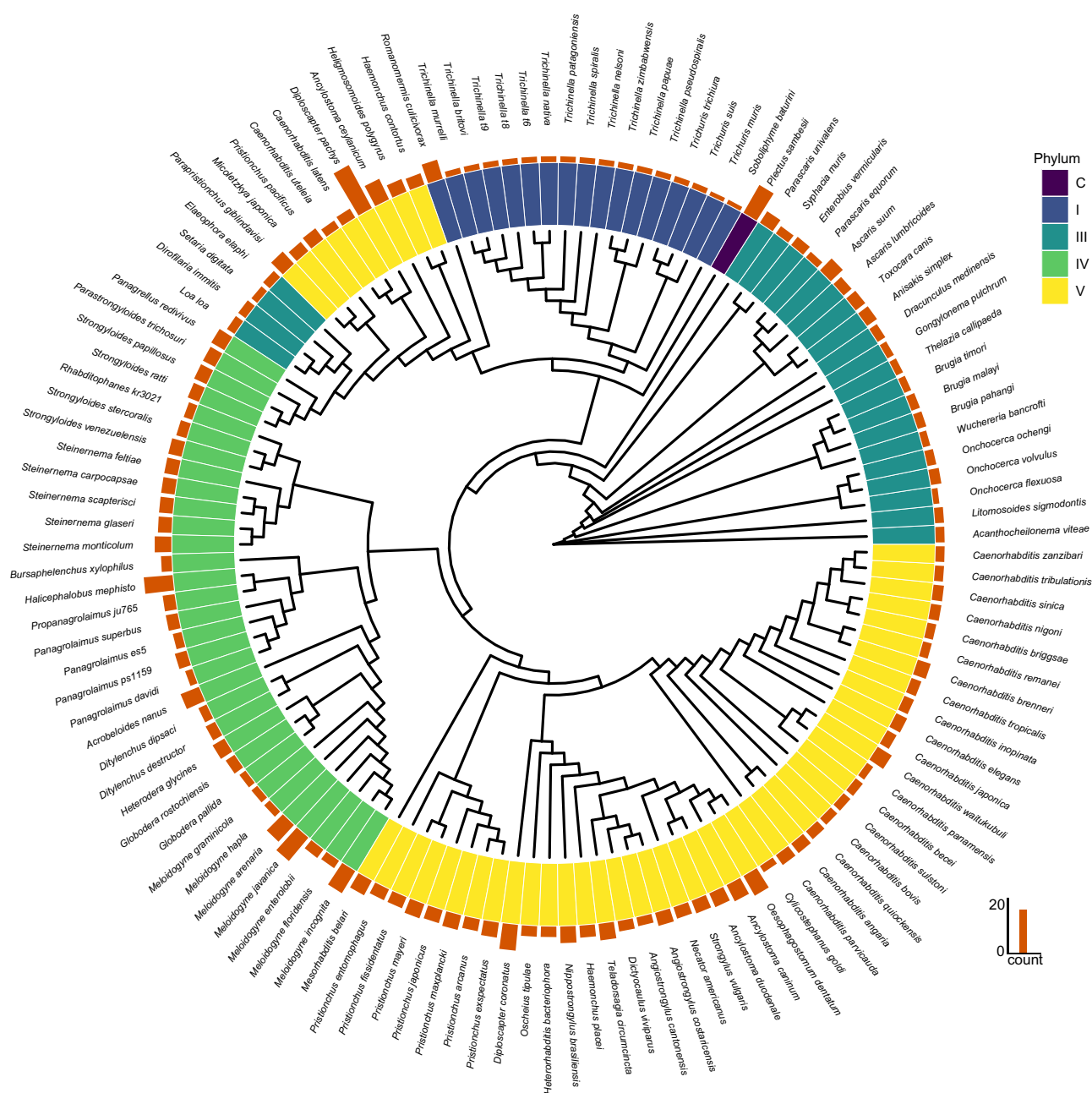
Genomes used in this study were obtained from WormBase ParaSite (ver. WBPS15) (Howe et al. 2017). GluCl subunits were predicted from each genome using hidden Markov models (HMMs) as described previously (Langeland et al. 2020, 2021; Wheeler et al. 2020; Bode and O'Halloran 2021). The Pfam (Finn et al. 2014) HMM, PF02932 (*Neur\_chan\_memb*), was used alongside an HMM based on known GluCl proteins that was generated using *hmmbuild* (Mistry et al. 2013). Next, *hmmsearch* (-tblout -noali) (Mistry et al. 2013) was used to compare HMMs against genomes obtained from the WormBase Parasite ftp site at the European Bioinformatics Institute (EBI: <http://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS15/species>). This approach yielded 983 predicted GluCl subunit genes across all 125 nematode species (Supplementary Table 1). From this starting dataset, alternative splice-forms were removed to get a more accurate gene count per species. After filtering, 864 GluCl subunit genes remained in the dataset and were used as queries in *blastp* (Camacho et al. 2009) searches against a database of characterized GluCl subunits from *C. elegans* and *H. contortus* so as to confirm GluCl identity. Predicted GluCl subunits were then aligned using MAFFT ver. 7.487 (Katoh et al. 2019). Phylogenetic relationships for the aligned orthologous clusters were then reconstructed using PhyML (Guindon et al. 2010) by employing an appropriate model of selection determined using ProtTest3 (Darriba et al. 2011). Heatmaps and trees were rendered and annotated using R (ver. 4.0.5) *heatmap* package (ver. 1.0.12) (Kolde 2019), *ggtree* package (ver. 2.4.2) (Yu et al. 2017), and *ggtreeExtra* (ver. 1.0.4) (Xu et al. 2021). Orthologous clusters that were refined using this phylogenetic approach were then used to make a database (*makeblastdb*) which was interrogated using local *blast* executables (Camacho et al. 2009) in order to assign orthology to divergent sequences as well as partial sequences from within our database that lacked a clear ortholog. Statistics were performed using Python (ver. 3.9.6) *scipy* module (ver. 1.7.0) and R ver. 4.1.1.

## Results and discussion

In order to predict GluCl subunit genes, our analyses included genomes from 125 nematode species (Fig. 1; Supplementary Table 1) that covered clades I (17 species), III (24 species), IV (32 species), V (51 species), and C (*Plectus sambesii*), and from this analysis, 864 GluCl subunit genes were identified (Fig. 1). To assess whether there were differences between GluCl subunit gene counts across different Clades, we normalized the predicted GluCl gene counts by dividing the total gene count per species by the number of species within each clade and then compared the normalized value with clade number (Supplementary Fig. 1). From this statistical test, we could conclude that our observed GluCl gene counts per clade were not different to what we would expect for each clade ( $P = 0.4918$  Chi-Square Goodness-of-fit test; note: this statistical test excluded clade C as it was represented by only 1 species, and instead only included clades I, III, IV, and V). A lack of discrepancy between our observed GluCl gene counts and expected counts does not rule out possible finer resolution differences, and in fact, we did observe differences in GluCl gene counts across individual species: the species with the most predicted GluCl genes were *Diploscapter pachys*, a free-living clade V nematode, and *P. sambesii* a free-living marine nematode within clade C, while the species with the least predicted GluCl genes were *Soboliphyme baturini*, a clade I intestinal parasite of mustelids and *Trichuris muris*, a clade I intestinal parasite of mice (Fig. 1).

Next, we tested whether genome completeness influenced GluCl gene counts. Benchmarking Universal Single-Copy Orthologs (BUSCO) scores are the standard measure of genome quality. BUSCO scores measure the completeness of an assembly by looking for the presence or absence of highly conserved genes. By comparing BUSCO scores for each species' genome assembly against predicted GluCl gene counts we did not observe a significant correlation (Supplementary Fig. 2:  $R = 0.0017$ ,  $P = 0.98$ ), suggesting that while genome improvement may alter some of the final GluCl gene counts, genome quality was not a major driver in the total GluCl gene count differences we observed. It is also worth noting that the majority of assemblies that we used to predict GluCl subunit genes have good BUSCO scores (Supplementary Fig. 3, median BUSCO score = 86.2).

To organize the 864 GluCl subunit genes that we identified across 125 species into orthologous groups, we performed *blastp* (Camacho et al. 2009) searches of all 864 sequences against a database of characterized GluCl subunits from *C. elegans* and *H. contortus*. This approach revealed clear orthologous clusters for 365 of the GluCl sequences which resulted in *blastp* e-values  $< 5 \times 10^{-324}$  (which are typically rounded to 0 via local *blastp* executables). The classification of these sequences was further tested by examining their phylogenetic relationship; all 365 sequences were aligned using MAFFT (Katoh et al. 2019) and from the alignment the best model of selection was determined to be VT (Müller and Vingron 2000) using ProtTest 3 (Darriba et al. 2011). This information was then used to infer the phylogeny for these conserved 365 GluCl sequences (Fig. 2). The phylogeny revealed that GLC-2, GLC-4, and AVR-14 form separate clades while GLC-1 and AVR-15 both group within the same clade and form a sister clade to GLC-3 (Fig. 2; Supplementary Table 2). These observations suggest that GLC-1 and AVR-15 may represent a more recent duplication event that may postdate a duplication event from GLC-3. This phylogenetic relationship is in keeping with recent observations from another group (Lamassiaude et al. 2021) and is further supported by the synteny of GluCl genes within *C. elegans* in which both GLC-1 and AVR-15 are closely linked on chromosome V while GLC-3 is located toward the other end of chromosome V (Supplementary Fig. 4). Within the conserved group of 365 GluCl sequences that we identified, there was notable expansions within specific species belonging to clades IV and V such as *Halicephalobus mephisto* which has an expansion of predicted *avr-14* subunit genes and also *D. pachys*, which has an expansion of predicted *glc-2* subunit genes (Fig. 3). The *Meloidogyne* species: *M. arenaria*, *M. javanica*, and *M. incognita* displayed expansions of *glc-3*, *glc-4*, and *avr-14* (Fig. 3). It is interesting that 2 species of free-living nematodes as well as 3 plant-parasitic nematode species had specific expansions, and this led us to test whether there were differences between nematode lifestyle (animal parasitic, entomopathogenic, free-living, human parasitic, or plant parasitic) and counts of the predicted GluCl subunit type (*glc-1*, *glc-2*, *glc-3*, *glc-4*, *avr-14*, or *avr-15*) from our conserved 365 sequences (Figs. 2 and 3). To test this, we normalized the gene counts for each classification group by the number of species within each lifestyle category (Supplementary Fig. 5), and then performed a Chi-Square Goodness-of-fit test as described above for our comparison between clade and GluCl gene counts. From this analysis, we did not observe a significant difference in our observed GluCl gene counts for each lifestyle from what would be expected ( $P = 0.8582$ ), suggesting that predicted GluCl gene expansions in species such as *H. mephisto* and *D. pachys* as well as the *Meloidogyne* species examined may reflect specific lifestyle adaptations.



**Fig. 1.** Radial tree representing the total predicted GluCl gene count per species (bar plot) and the corresponding phylum (interior color annotation). The phylogenetic relationship between each species was based on the *rpb-1* gene. Alignments were made using MAFFT (Kato et al. 2019) while ProtTest 3 (Darriba et al. 2011) was used to determine the best model of selection, which was implemented within PhyML (Guindon et al. 2010) to reconstruct the phylogenetic relationship. Rendering and annotation of the tree was done using the *ggtree* (Yu et al. 2017) and *ggtreeExtra* (ver. 1.0.4) (Xu et al. 2021) packages.

To classify the remaining 499 GluCl subunit genes from the 864 total GluCl gene predictions into orthologous clusters, we used the conserved 365 predicted GluCl subunit genes that we described above alongside functionally characterized GluCl subunits (Lamassiaude et al. 2021) to create a database which we interrogated using *blastp* searches to cluster into orthologous groups based on top *blastp* hits. These data alongside the conserved 365 predicted GluCl subunit genes were converted into JSON tables to build an online database that included all 864 predicted GluCl sequences. The database is available at the following URL: [http://ohalloranlab.net/nematode\\_GluCl](http://ohalloranlab.net/nematode_GluCl) and includes the orthology evidence for each entry (i.e. *phylogeny* for the 365

conserved GluCl subunit genes and *Blast* for the 499 divergent GluCl subunit genes)

The more divergent 499 predicted GluCl sequences that we identified lacked significant conservation across all the sequences which made model selection and phylogenetic reconstructions difficult to interpret. Some of these more divergent predicted GluCl genes exhibited sequence diversity within putative IVM binding sites. While Glu binds the extracellular N-terminal region of GluCl channels, IVM can penetrate to the TM domains and interact with a number of key residues (Hibbs and Gouaux 2011; Althoff et al. 2014). We identified substitutions at a key Glycine residue within TM3 which

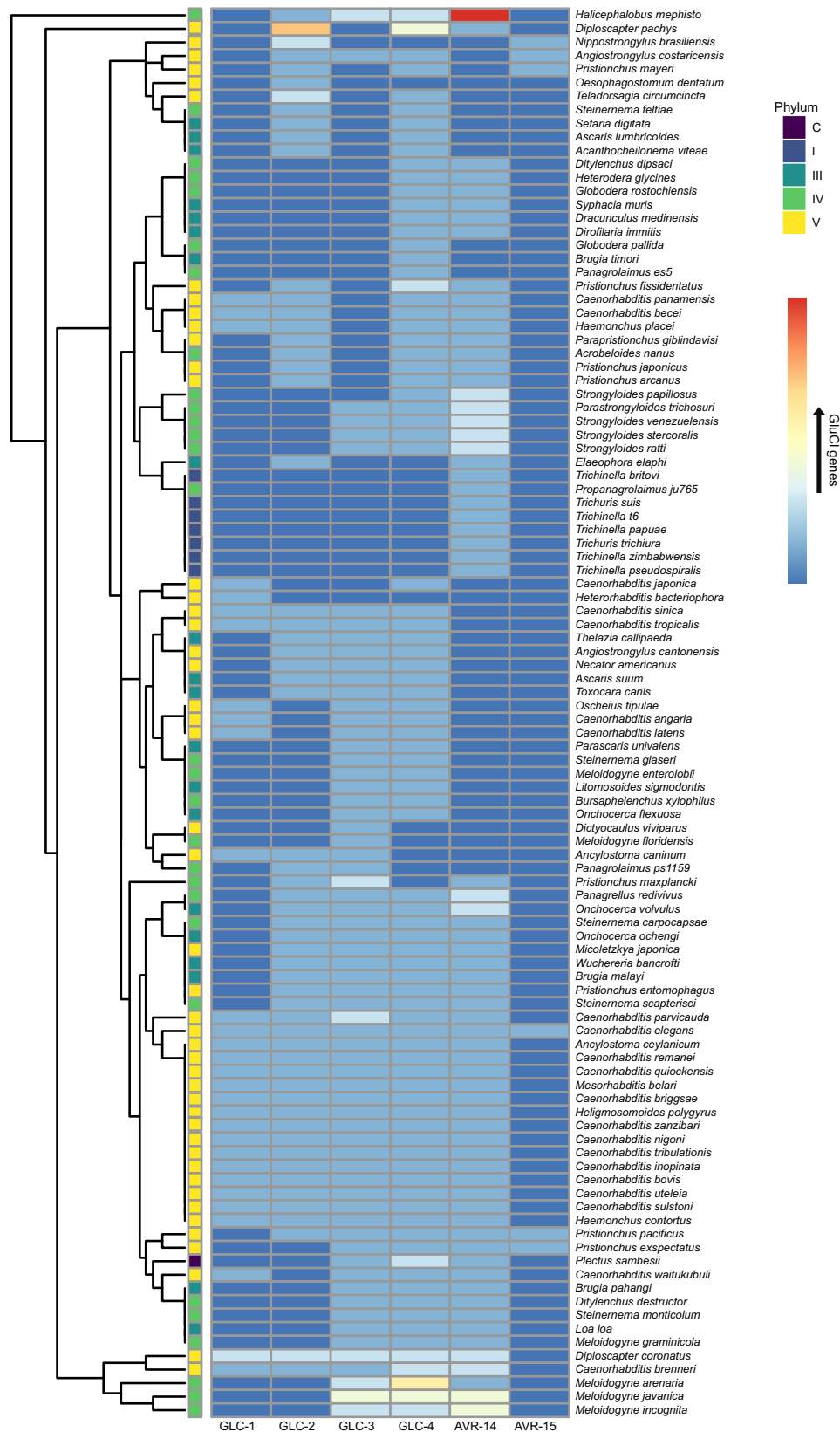


**Fig. 2.** Phylogeny of conserved clusters of predicted GluCl genes inferred using PhyML (Guindon et al. 2010). Sequences clustered into groups of GLC-2, GLC-4, and AVR-14 while GLC-1 and AVR-15 clustered together forming a sister clade to GLC-3.

was mutated to an Alanine within some of the divergent GluCl gene sequences we identified (Supplementary Fig. 6). This Glycine residue is referred to as M3-Gly (Lynagh and Lynch 2010), and has been studied in other species to characterize the effect on drug binding. In *H. contortus*, wildtype GluClR  $\alpha$ 3B exhibited  $EC_{50} = 39 \pm 6$  nM, while mutation of M3-Gly to Serine resulted in  $EC_{50} = 620 \pm 140$  nM, and mutation of M3-Gly to Alanine caused  $EC_{50} = 1.2 \pm 0.3$   $\mu$ M (Lynagh and Lynch 2010). Interestingly, these mutations also altered Glu sensitivity from wildtype  $EC_{50} = 15.3 \pm 1.8$   $\mu$ M to  $EC_{50} = 108 \pm 10$   $\mu$ M when M3-Gly is mutated to a Serine and  $EC_{50} = 154 \pm 21$   $\mu$ M when M3-Gly is mutated to Alanine (Lynagh and Lynch 2010).

The data described in this paper is available as a searchable database at the following URL: [http://ohalloranlab.net/nematode\\_GluCl](http://ohalloranlab.net/nematode_GluCl). The search textbox at the top of the landing page

can be used to query the database for identifiers such as species or GluCl subunit while the columns can be sorted by the clicking the column headings. The raw protein sequence for each entry can be viewed by clicking the green circle in the first cell of a given row, and each entry also links out to WormBase Parasite (Howe et al. 2017) for further exploration. The database can also be searched for similar sequences using local blast executables (Camacho et al. 2009) by simply entering a protein sequence into the input field and clicking the submit button to interrogate the nematode GluCl database. We plan to include more species to the database as they become available, and view this database as a platform for users to ask questions related to GluCl evolution and function across diverse nematode species while also providing molecular insights into GluCl pharmacology.



**Fig. 3.** Heatmap of conserved GluCl subunit genes from Fig. 2. Each row corresponds to an individual species while the vertical annotations denote phylum for each species. Hot colors indicate increased predicted GluCl counts while cooler colors represent lower predicted GluCl gene counts.

## Data availability

The data underlying this article are available in the article, in its online supplementary material, and at the following URL: [http://ohalloranlab.net/nematode\\_GluCl](http://ohalloranlab.net/nematode_GluCl).

[Supplemental material](#) is available at G3 online.

## Acknowledgments

Support for this research was provided by the George Washington University (GWU) Columbian College of Arts and Sciences and Department of Biological Sciences.

## Funding

The funding for this research was provided by the George Washington University Columbian College of Arts and Sciences and Department of Biological Sciences.

## Conflicts of interest statement

None declared.

## Literature cited

- Althoff T, Hibbs RE, Banerjee S, Gouaux E. X-ray structures of GluCl in apo states reveal a gating mechanism of Cys-loop receptors. *Nature*. 2014;512(7514):333–337.
- Arena JP, Liu KK, Paress PS, Frazier EG, Cully DF, Mrozik H, Schaeffer JM. The mechanism of action of avermectins in *Caenorhabditis elegans*: correlation between activation of glutamate-sensitive chloride current, membrane binding, and biological activity. *J Parasitol*. 1995;81(2):286–294.
- Bode K, O'Halloran DM. Pan-phylum genome-wide identification of sodium calcium exchangers reveal heterogeneous expansions and possible roles in nematode parasitism. *Gene*. 2021;810:146052.
- Calimet N, Simoes M, Changeux J-P, Karplus M, Taly A, Cecchini M. A gating mechanism of pentameric ligand-gated ion channels. *Proc Natl Acad Sci U S A*. 2013;110(42):E3987–E3996.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
- Cully DF, Vassilatis DK, Liu KK, Paress PS, Van der Ploeg LH, Schaeffer JM, Arena JP. Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. *Nature*. 1994;371(6499):707–711.
- Daeffler KN-M, Lester HA, Dougherty DA. Functional evaluation of key interactions evident in the structure of the eukaryotic Cys-loop receptor GluCl. *ACS Chem Biol*. 2014;9(10):2283–2290.
- Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011;27(8):1164–1165.
- Dent JA, Davis MW, Avery L. avr-15 encodes a chloride channel subunit that mediates inhibitory glutamatergic neurotransmission and ivermectin sensitivity in *Caenorhabditis elegans*. *EMBO J*. 1997;16(19):5867–5879.
- Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014;42(Database issue):D222–D230.
- Ghosh R, Andersen EC, Shapiro JA, Gerke JP, Kruglyak L. Natural variation in a chloride channel subunit confers avermectin resistance in *C. elegans*. *Science*. 2012;335(6068):574–578.
- Guindon S, Dufayard J-F, 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(3):307–321.
- Hibbs RE, Gouaux E. Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature*. 2011;474(7349):54–60.
- Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite—a comprehensive resource for helminth genomics. *Mol Biochem Parasitol*. 2017;215:2–10.
- Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2019;20(4):1160–1166.
- Kolde R. Pheatmap: pretty heatmaps package. 2019. <https://rdrr.io/cran/pheatmap>.
- Lamassiaude N, Courtot E, Corset A, Charvet CL, Neveu C. Pharmacological characterization of novel heteromeric GluCl subtypes from *C. elegans* and parasitic nematodes. *Br J Pharmacol*. 2021. <https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bph.15703>.
- Langeland A, Hawdon JM, O'Halloran DM. NemChR-DB: a database of parasitic nematode chemosensory G-protein coupled receptors. *Int J Parasitol*. 2020;51(5):333–337.
- Langeland A, Jetter H, O'Halloran DM. The diversity of ABC transporter genes across the Phylum Nematoda. *Parasitol Int*. 2021;83:102357.
- Lynagh T, Lynch JW. A glycine residue essential for high ivermectin sensitivity in Cys-loop ion channel receptors. *Int J Parasitol*. 2010;40(13):1477–1481.
- McCavera S, Rogers AT, Yates DM, Woods DJ, Wolstenholme AJ. An ivermectin-sensitive glutamate-gated chloride channel from the parasitic nematode *Haemonchus contortus*. *Mol Pharmacol*. 2009;75(6):1347–1355.
- Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*. 2013;41(12):e121.
- Müller T, Vingron M. Modeling amino acid replacement. *J Comput Biol*. 2000;7(6):761–776.
- Vassilatis DK, Arena JP, Plasterk RH, Wilkinson HA, Schaeffer JM, Cully DF, Van der Ploeg LH. Genetic and biochemical evidence for a novel avermectin-sensitive chloride channel in *Caenorhabditis elegans*. Isolation and characterization. *J Biol Chem*. 1997;272(52):33167–33174.
- Wheeler NJ, Heimark ZW, Airs PM, Mann A, Bartholomay LC, Zamanian M. Genetic and functional diversification of chemosensory pathway receptors in mosquito-borne filarial nematodes. *PLoS Biol*. 2020;18(6):e3000723.
- van Wyk JA, Malan FS. Resistance of field strains of *Haemonchus contortus* to ivermectin, closantel, rafoxanide and the benzimidazoles in South Africa. *Vet Rec*. 1988;123(9):226–228.
- Xu S, Dai Z, Guo P, Fu X, Liu S. ggtreeExtra: compact visualization of richly annotated phylogenetic data. *Mol Biol Evol*. 2021;38(9):4039–4042.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 2017;8(1):28–36.