PLoS one

Domain-Based Identification and Analysis of Glutamate Receptor Ion Channels and Their Relatives in Prokaryotes

Mao-Feng Ger[1], Gloria Rendon[2], Jeffrey L. Tilson[3], Eric Jakobsson[1,2]*

1 Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, 2 National Center for Supercomputing Applications, Urbana, Illinois, United States of America, 3 Renaissance Computing Institute, Raleigh, North Carolina, United States of America

Abstract

Voltage-gated and ligand-gated ion channels are used in eukaryotic organisms for the purpose of electrochemical signaling. There are prokaryotic homologues to major eukaryotic channels of these sorts, including voltage-gated sodium, potassium, and calcium channels, Ach-receptor and glutamate-receptor channels. The prokaryotic homologues have been less well characterized functionally than their eukaryotic counterparts. In this study we identify likely prokaryotic functional counterparts of eukaryotic glutamate receptor channels by comprehensive analysis of the prokaryotic sequences in the context of known functional domains present in the eukaryotic members of this family. In particular, we searched the nonredundant protein database for all proteins containing the following motif: the two sections of the extracellular glutamate binding domain flanking two transmembrane helices. We discovered 100 prokaryotic sequences containing this motif, with a wide variety of functional annotations. Two groups within this family have the same topology as eukaryotic glutamate receptor channels. Group 1 has a potassium-like selectivity filter. Group 2 is most closely related to eukaryotic glutamate receptor channels. We present analysis of the functional domain architecture for the group of 100, a putative phylogenetic tree, comparison of the protein phylogeny with the corresponding species phylogeny, consideration of the distribution of these proteins among classes of prokaryotes, and orthologous relationships between prokaryotic and human glutamate receptor channels. We introduce a construct called the Evolutionary Domain Network, which represents a putative pathway of domain rearrangements underlying the domain composition of present channels. We believe that scientists interested in ion channels in general, and ligand-gated ion channels in particular, will be interested in this work. The work should also be of interest to bioinformatics researchers who are interested in the use of functional domain-based analysis in evolutionary and functional discovery.

Introduction

It is estimated that 20%–40% of genes code for integral membrane proteins in archaea, bacteria, and eukaryote [1]. Because of the enormous energy barrier associated with moving ions across lipid bilayers [2] (Figure 1), proteins are essential for the transmembrane movement of polar and charged substances. Specific transmembrane proteins, like ion channels, transporters and pumps, appear to have arisen in very early forms of cellular life [3].

Ion channels are specialized transmembrane proteins through which cations or anions move passively down the electrochemical gradients that are created by ion pumps. Ion channels differ greatly in their structural and functional properties and are classified by their selectivity (Na^+, K^+, Ca^{2+} and Cl^-) and activation mechanism (voltage-gated or ligand-gated). The largest subfamily of ion channels is comprised of the pore-loop channels, all of which carry a basic structural unit – a re-entrant pore-loop flanked by two transmembrane helices (TM's). (Figure 2) The ion selectivity is conferred by the pore-loop [4]. This common topology can be interpreted to suggest that the pore-loop channels have a common ancestor. This suggestion was born out by the

discovery of a prokaryotic channel that contained the ligand-binding extracellular domain characteristic of glutamate receptor channels but a pore-loop characteristic of a potassium channel [5].

Glutamate, a major excitatory neurotransmitter, activates two receptor families: metabotropic glutamate receptor proteins (mGluR), which activate biochemical cascades, and ionotropic glutamate receptors, which form cation selective ion channels (iGluR) and are members of the pore-loop subfamily. Compared to the voltage-gated members of the pore-loop subfamily, iGluR's have opposite transmembrane orientation to the others (the pore-loop re-enters from the intracellular side). There are three major eukaryotic iGluR's subtypes, the AMPA, kainite and NMDA receptors, which form cation channels permeable to Na^+, K^+ and Ca^{2+}. Because of the difficulty of purification and crystallization of integral membrane proteins, we only have the high resolution structure for the extracellular ligand-binding domain of iGluR [6]. Some critical amino acids are identified in ligand-binding sequence.

In addition to the above-mentioned glutamate-receptor channel homologue, many other homologues to mammalian ion channels have been found in sequenced prokaryotic genomes, such as K^+ channels, Na^+ channels, and Cl^- channels [7]. In addition Kuner,
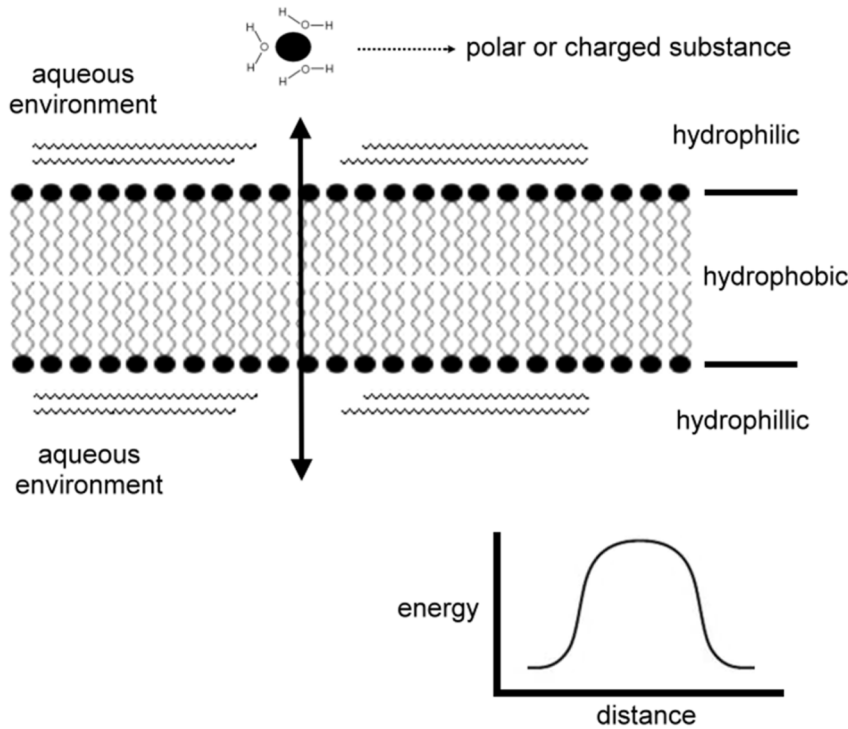
Figure 1. Energy barrier for an ion to move across the membrane.
doi:10.1371/journal.pone.0012827.g001

et al [8] noted the existence of other prokaryotic sequences bearing a resemblance to eukaryotic glutamate receptor channels.

The relative simplicity of prokaryotic ion channels makes them excellent objects for biophysical research [9]. A particularly notable example is the use of a prokaryotic potassium channel to make the first high resolution structure determination of voltage gated channels [10]. In many ways studying prokaryotic

homologues can shed significant light on eukaryotic channels, as well the prokaryotic channels being of interest in their own right. For these reasons, a few years ago our laboratory (in collaboration with the laboratory of I. Aravind at NIH) set out to find prokaryotic homologues to the Ach receptor channel family. A straightforward BLAST [11] search yielded no results. We therefore undertook a search based on finding sequences with
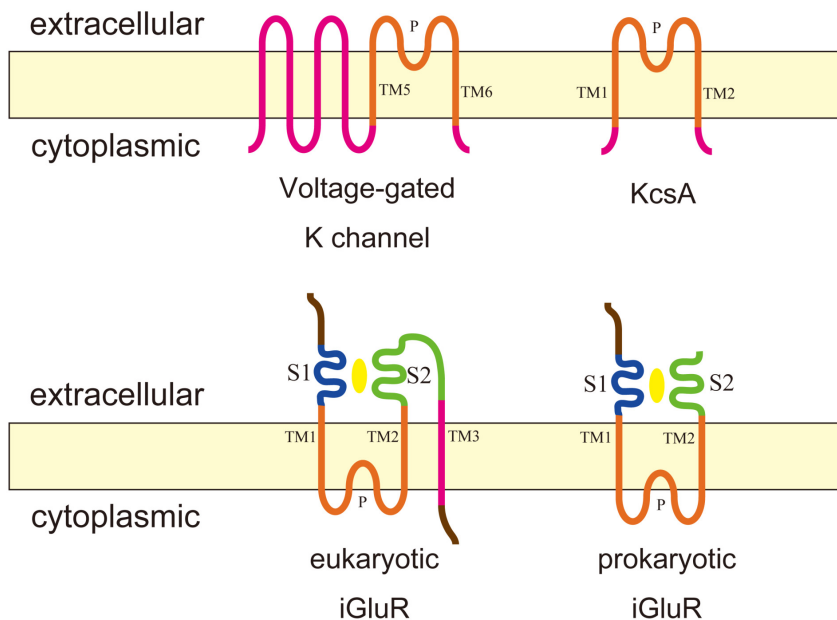
Figure 2. Topology diagrams of pore-loop ion channels.
doi:10.1371/journal.pone.0012827.g002

conserved domains characteristic of Ach receptor channel proteins and with the appropriate topology. That approach yielded a number of predicted prokaryotic members of this channel family [12]. One of our predicted channes was cloned, expressed, and functionally characterized as a channel [13] and high resolution structures were determined [14]. We anticipate that comprehensive identification of members of this group will lead to further functional and structural characterization of this family of channels, as well as insights into evolutionary and comparative aspects of channel biology. In the present study we extend this approach to a systematic domain-based search to identify and characterize in the nonredundant protein database all the prokaryotic homologues of the glutamate receptor channel family; i.e., prokaryotic iGluR's.

Materials and Methods

Searching for Prokaryotic iGluR's

The overall strategy for discovery of the prokaryotic iGluR's is provided in the flow chart of the five stage screening process, plus a validation stage using the InterPro database, in Figure 3(a).

We begin the search with the sequence iGluR0 from Synechocystic PCC6803 [15] which has been well characterized both functionally [5] and structurally [16]. At stage 1 in Figure 3a, we used PSI-BLAST [11] to search the SDSC nonredundant protein database for the S1 binding region (NSEYVRQNSISAGITAVAEGELDILIGPISVTPERAAIEGITFTQPYFSSGIGLLIP, 57 aa long). This returned 2314 sequences with an E-value below 10. We applied the same method separately with the S2 segment of the binding region (EAVMFDRPALIYYTRQNPNLNLEVTEIRVSLEPYGFVLKENSPLQKTINVEMLNLLYSRVIAEFTERWL, 69 aa long) and returned 2344 sequences. At stage 2 in Figure 3a, we invoked TransMembrane Hidden Markov Model [TMHMM] [17] to predict the number of transmembrane (TM) helices in each sequence. We eliminated all sequences with fewer than 2 TM's, which is the minimal number for the iGluR structure. This left us with 758 sequences with S1 and at least 2 TM's and with 731 sequences with S2 and at least 2 TM's. At stage 3, we separated the prokaryotic sequences from the eukaryotes. We found 135 sequences with S1 and 2 TM's and 132 sequences with S2 and 2 TM's. At stage 4, out of the 135 and the 132 we keep only the sequences that have both S1 and S2, which total 100 (see Data S1 for detailed list). The annotations of the 100 sequences, clearly related to each other, are varied. In the definition line of the SDSC nonredundant protein database, 51 of them are annotated as ABC-type amino acid transporter or something similar, 13 of them are annotated as binding proteins, 14 of them are annotated as hypothetical proteins, 2 of them are annotated as K channels, plus some other scattered annotations (Table 1).

To explore the relationships among the 100 sequences, we aligned the sequences with ClustalW [18] and built a phylogenetic tree for them by DRAWGRAM [19]. The result is shown in Figure 4.

A notable feature of Figure 4 is that in many cases there is a disconnect between how close the sequences are on the tree and the similarity of the annotations. In some cases proteins that are quite similar are annotated differently, while sequences that seem quite far apart have the same annotation. A BLAST [11] of each of the 100 was done against the nonredundant database (data not shown) and confirmed that the sequence that gave the best hit was usually the one that was closest on the tree, and that the closest one on the tree was always one of the top few.

We then performed a topology analysis (stage 5 in Figure 3a) for the 100 sequences. The transmembrane regions are determined by TMHMM [17] and the glutamate binding regions are determined by sequence alignment. Through the visualization tool SeqVISTA [20], we can see the relative positions and lengths for TM's and glutamate binding regions in each protein. 22 of the 100 can be identified as having the characteristic topology of glutamate receptor channels; i.e., the S1 and S2 glutamate binding domains flanking two TM helices (M1 and M2 region), in turn flanking a pore-loop (a domain that looks like a partial TM helix, P region). (One of the 22 sequences is the authoritative sequence that we used as our initial probe [5].) Figure 3b shows the e-values and TM probability scores for the S1/S2 and TM regions of the 22 sequences. It is seen that the statistical evidence for the identification and the topology are very strong. Figure 5 shows the SeqVISTA pattern characteristic of these 22 sequences and, for comparison, the SeqVISTA pattern for the human glutamate receptor channel orthologous (by the standard of reciprocal best hits) to the particular prokaryotic sequence shown. There are some differences. The human proteins are much larger, having an extra TM near the C-terminus. But there is a major similarity, i.e., the glutamate binding domains flanking two TM domains and a pore-loop. The supplementary material (Data S3) includes the SeqVISTA diagrams for all 100 prokaryotic sequences in our search. Besides the 22 sequences, the other 78 prokaryotic sequences that have the glutamate binding domain and two or more TM helices have somewhat different topologies.

Results

Features and Evolution of the Prokaryotic Glutamate Receptor Channels

Of the 22 putative channels, 12 of them have a distinctive potassium channel selectivity filter. We designate these as our Group 1. The other 10 have P regions we do not recognize as distinctively similar to any channel with a known particular selectivity. Their annotations in the SDSC nonredundant protein database are shown in Table 2. Based on our analysis we would suggest that Group 1 be annotated as "putative glutamate-sensitive potassium channel" (except for #56, for which the word "putative" should be left off, since it has been functionally characterized as a glutamate-sensitive potassium channel [5].) We would suggest that Group 2 be annotated as "putative glutamate-sensitive ion channel". Besides TM, we also used signalP [21] to test the existence of signal peptide. We found that two members of Group1 and two members of Group 2 lack the signal peptides which help the orientation of ion channel. The reasons for this may be the following: 1) They are pseudogenes; 2) they may have a different mechanism of inserting into membranes, or 3) they are oppositely oriented in the membrane than the other Group 1 and Group 2 channels. Motif searching has important significance in predicting the structures and functions of proteins. Therefore, we analyze the protein sequences by InterProScan [22] which is a web-based motif searching tool (http://www.ebi.ac.uk/interpro/) and federates 13 InterPro member databases into one resource. By searching the different protein signature databases, we can get a more comprehensive understanding of our target proteins. In order to efficiently utilize InterProScan, we developed a high throughput workflow around the InterProScan core program, that we call MotifNetwork [23].

Through MotifNetwork, we found that all 100 sequences have a glutamate binding motif, which was expected because we took glutamate binding region as our PSI-BLAST probe. We also found that none of the Group 1 or Group 2 members had a domain characteristic of ABC transporters, reinforcing our view,

nonredundant protein database

7 rounds of PSI-BLAST by S1 of iGluR0

e-value ≤ 10

stage 1

nonredundant protein database

7 rounds of PSI-BLAST by S2 of iGluR0

2314 sequences

pass the TM probability test

stage 2

2344 sequences

TMHMM analysis: ≥ 2 transmembranes

TMHMM analysis: ≥ 2 transmembranes

758 sequences

from Prokaryotes

stage 3

731 sequences

135 sequences

have both S1 and S2

stage 4

132 sequences

100 sequences

topology ananlysis: S1/S2 flank 2 or 3 transmembrane

stage 5

12 sequences of Group 1
10 sequences of Group 2

InterPro domain ananlysis with identical/similar domain composition

validation stage

12 sequences of Group 1
10 sequences of Group 2

(a)

(b)

#	E-value for S1	E-value for S2	TM probability for M1	TM probability for P	TM probability for M2
Group 1					
23	3.00E-07	1.00E-09	0.99999	0.89394	0.99275
25	9.00E-08	2.00E-09	1	0.99337	0.99813
33	0.018	8.00E-09	0.99999	0.73089	0.99722
37	1.00E-08	3.00E-08	1	0.99092	0.99848
40	1.00E-07	4.00E-08	0.99996	0.96695	0.99982
41	6.00E-05	6.00E-08	0.99997	0.97498	0.99693
43	1.00E-06	7.00E-08	1	0.69016	0.99812
46	3.00E-07	1.00E-07	0.99975	0.70932	0.99858
56	6.00E-10	6.00E-07	0.99984	0.06564	0.99552
62	2.00E-10	1.00E-06	0.99993	0.88389	0.99434
65	6.00E-04	3.00E-06	1	0.77978	0.99838
94	2.00E-05	2.1	1	0.43197	0.99847
Group 2					
1	5.00E-07	6.00E-16	0.99999	0.45822	0.99995
2	2.00E-06	7.00E-16	0.99998	0.47037	0.99995
3	1.00E-06	9.00E-16	0.99999	0.47272	0.99995
4	1.00E-06	1.00E-15	0.99998	0.3648	0.99995
5	2.00E-08	1.00E-15	0.99998	0.31253	0.99836
6	2.00E-06	1.00E-13	0.99924	0.95602	0.99996
9	3.00E-06	9.00E-13	0.99999	0.19217	0.99835
42	1.00E-03	6.00E-08	0.99988	0.78611	0.99969
50	2.00E-04	2.00E-07	0.99999	0.95594	0.99994
63	4.00E-07	1.00E-06	0.9989	0.02518	0.99825

Figure 3. The searching strategy for finding prokaryotic iGluR's and the statistical proof. The strategy in (a) includes 5 stages and an additional validation stage. At each stage, we select protein sequences which are qualified for the requirements. In (b), the statistical e-values for S1 and S2 identification and TM probability scores by PSI-BLAST and TMHMM, respectively. The TM probability scores which do not pass the TM probability test are shaded in (b) and not counted as TM'S.
doi:10.1371/journal.pone.0012827.g003

stated above, that such annotation for those particular sequences is in error.

The results of the above are summarized in an Evolutionary Domain Network (EDN) (Figure 6). In the EDN representation, the proteins are grouped into domain sets according to the domain composition of each. (By "domain composistion" we mean the list of domains contained in the set.) The first row of the EDN contains all domain sets that consist of only a single domain. The second row contains those domain sets with two domains, the third row with three, etc. Tie lines are drawn between domain sets that

Table 1. Annotation of 100 bacterial sequences found to contain glutamate binding domains and two transmembrane domains.

gene annotation	protein No.	quantity
ABC transport system glutamine-binding protein	1,5,7,15,18,31,36,39,53,58,59,61	12
ABC-type amino acid transport/signal	2,4,6,8,10,12,13,17,24,26,27,28,29,35,38,45,47,48, 49,54,55,64,69,71,77, 78,81,82,83,85,88,90,91,95, 96,97,98, 99,52	39
transporter	19,21	2
binding protein	66,80	2
extracellular solute-binding protein	9,25,33,40,41,46,63,65,70,73,89	11
hypothetical protein	16,20,22,43,50,56,57,60,67,72,75,76, 79,100	14
iGluR	3,23,37,62	4
K channel	42,94	2
sensory transduction protein kinase	34	1
sensory box protein	87	1
IMP dehydrogenase/GMP reductase	84	1
Unknown function	30,32,44,51,68,74,86,92,93	9

can be derived from each other by the addition or subtraction of a single domain, representing roughly the evolutionary process of domain recombination. It should be noted that we have not screened out overlapping domains. Thus in some cases the same section of the protein sequence may be represented by two domain designations. We did attempt to screen overlaps, but any automated overlap screening resulted in loss of significant information, so we elected to report all MotifNetwork hits regardless of overlap.

By inspection of Figure 6, we see that all Group 1 sequences contain the IPR013099, whose short title is Ion transport 2. This domain represents a K$^+$ channel selectivity filter. As far as we have been able to determine so far, the combination of glutamate channel binding site and potassium channel selectivity filter represented by Group 1 is only in bacteria. No members of Group 1 can be found in archaea, neither can Group 2.

All Group 2 sequences have two domains in common: IPR001638 (Bacterial extracellular binding protein) and IPR015638 (glutamate receptor related). These are overlapping regions. The selectivity filter and permeation pathway have not apparently been defined as a distinctive InterPro domain.

Just one domain set appears disconnected from the others, and is placed on the right hand side of Figure 6. This contains domains IPR000515 and IPR013099. Only one protein (#94) is contained in this domain set. The existence of the potassium channel selectivity filter, plus the orientation of the glutamate binding domains to the transmembrane domains, defines this as a Group 1 channel. However the domain IPR000515, with this one exception, is only associated with the other sequences that do not have the structure of the glutamate binding domains flanking two TM domains and a pore-loop. It thus appears that sequence 94, despite its outlier status in Figure 6, may be a part of a linkage between the channel proteins and the non-channel proteins in this study. The intermediate domain sets have either vanished or have not yet been sequenced.

Inspection of Figure 6 shows that Human iGluR's can be connected to the prokaryotic scheme by intermediate steps equivalent to the net exchange of IPR001508 with IPR0016308 between NMDA receptor channels and Group 2 prokaryotic channels. This implies that Group 2 proteins might share a closer relationship to eukaryotic iGluR's than other prokaryotic

glutamate-binding proteins and NMDA's are closer to prokaryotic iGluR's than are other eukaryotic iGluR's. Delta 1 protein reacquired IPR001638 (otherwise only found in prokaryotes among the group we are studying) in its motif composition, which may result from a genetic recombination from outside (for example virus-mediated transfer from prokaryortes). It may be that some of the missing intermediates will appear in a more complete study of all the eukaryotic members of this family, which will be the subject of a future study.

Sequence analysis of Group 1 and Group 2 sequences

In order to identify the possible functions of Group 1 and Group 2 prokaryotic genes, we first made a multiple sequence alignment. In order to optimize the alignment, we align the domains separately and then join the alignments. We used the domain definitions of Mayer et al. [16] for the S1, S2, and channel regions (M1, P and M2). The conservation comparison is listed as Table 3. We can see that Group 2 is more conserved in glutamate binding region than Group 1 but less conserved in channel region.

In previous research about prokaryotic iGluR, scientists have identified some amino acids which are important in glutamate binding [5], specifically an Arg in S1 which interacts with α–carboxy group of L-glutamate and an Asp in S2 which interacts with α–amino group of L-glutamate. These are totally conserved in the Group 1 and Group 2 alignments. This conservation is shown in Figure 7. (The full alignments are shown in Data S4).

Phylogenetic analysis of Group 1 and Group 2 sequences

We made phylogenetic trees for the different regions (S1, S2, and P region) in Group 1 and Group 2 sequences. (Data S5) It is seen that the trees have essentially the same structure. We can conclude that the glutamate binding region and channel region have remained together for a long time in evolutionary history.

We compared the phylogenetic tree of 16s rRNA genes with the phylogenetic tree of Group 1 and Group 2 genes in Figure 8. In this figure it is seen that in the tree of protein sequences (right hand tree) the Group 1 sequences (red) are clearly clustered together and separate from the Group 2 sequences (green). However in the 16s RNA sequences, the organisms containing Group 1 and Group 2 do not separate into distinct clusters from each other, indicating
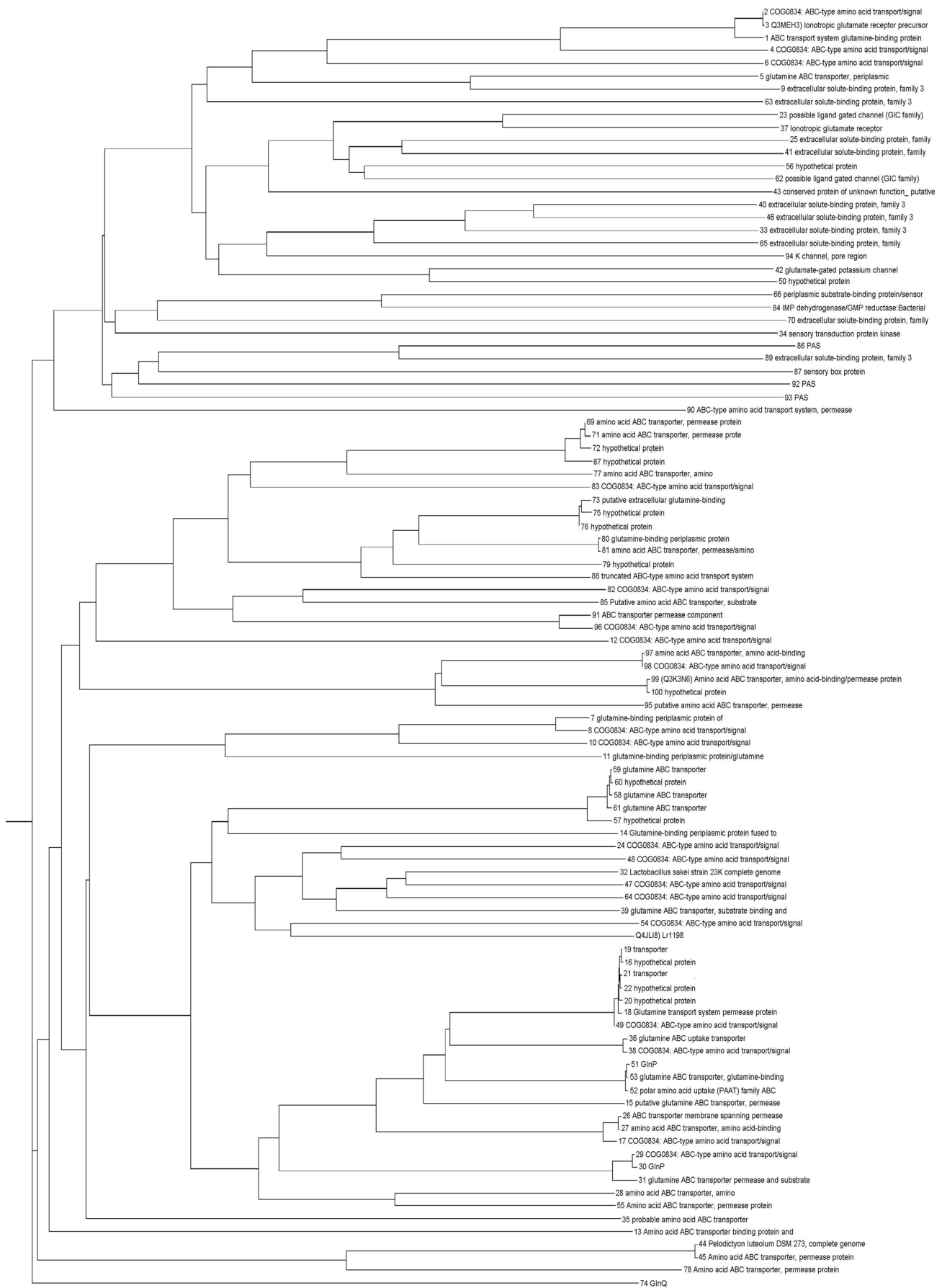
2 COG0834: ABC-type amino acid transport/signal
3 Q3MEH3) Ionotropic glutamate receptor precursor
1 ABC transport system glutamine-binding protein
4 COG0834: ABC-type amino acid transport/signal
6 COG0834: ABC-type amino acid transport/signal
5 glutamine ABC transporter, periplasmic
9 extracellular solute-binding protein, family 3
63 extracellular solute-binding protein, family 3
23 possible ligand gated channel (GIC family)
37 Ionotropic glutamate receptor
25 extracellular solute-binding protein, family
41 extracellular solute-binding protein, family
56 hypothetical protein
62 possible ligand gated channel (GIC family)
43 conserved protein of unknown function_ putative
40 extracellular solute-binding protein, family 3
46 extracellular solute-binding protein, family 3
33 extracellular solute-binding protein, family 3
65 extracellular solute-binding protein, family
94 K channel, pore region
42 glutamate-gated potassium channel
50 hypothetical protein
66 periplasmic substrate-binding protein/sensor
84 IMP dehydrogenase/GMP reductase:Bacterial
70 extracellular solute-binding protein, family
34 sensory transduction protein kinase
86 PAS
89 extracellular solute-binding protein, family 3
87 sensory box protein
92 PAS
93 PAS
90 ABC-type amino acid transport system, permease
69 amino acid ABC transporter, permease protein
71 amino acid ABC transporter, permease prote
72 hypothetical protein
67 hypothetical protein
77 amino acid ABC transporter, amino
83 COG0834: ABC-type amino acid transport/signal
73 putative extracellular glutamine-binding
75 hypothetical protein
76 hypothetical protein
80 glutamine-binding periplasmic protein
81 amino acid ABC transporter, permease/amino
79 hypothetical protein
88 truncated ABC-type amino acid transport system
82 COG0834: ABC-type amino acid transport/signal
85 Putative amino acid ABC transporter, substrate
91 ABC transporter permease component
96 COG0834: ABC-type amino acid transport/signal
12 COG0834: ABC-type amino acid transport/signal
97 amino acid ABC transporter, amino acid-binding
98 COG0834: ABC-type amino acid transport/signal
99 (Q3K3N6) Amino acid ABC transporter, amino acid-binding/permease protein
100 hypothetical protein
95 putative amino acid ABC transporter, permease
7 glutamine-binding periplasmic protein of
8 COG0834: ABC-type amino acid transport/signal
10 COG0834: ABC-type amino acid transport/signal
11 glutamine-binding periplasmic protein/glutamine
59 glutamine ABC transporter
60 hypothetical protein
58 glutamine ABC transporter
61 glutamine ABC transporter
57 hypothetical protein
14 Glutamine-binding periplasmic protein fused to
24 COG0834: ABC-type amino acid transport/signal
48 COG0834: ABC-type amino acid transport/signal
32 Lactobacillus sakei strain 23K complete genome
47 COG0834: ABC-type amino acid transport/signal
64 COG0834: ABC-type amino acid transport/signal
39 glutamine ABC transporter, substrate binding and
54 COG0834: ABC-type amino acid transport/signal
Q4JLI8) Lr1198
19 transporter
16 hypothetical protein
21 transporter
22 hypothetical protein
20 hypothetical protein
18 Glutamine transport system permease protein
49 COG0834: ABC-type amino acid transport/signal
36 glutamine ABC uptake transporter
38 COG0834: ABC-type amino acid transport/signal
51 GlnP
53 glutamine ABC transporter, glutamine-binding
52 polar amino acid uptake (PAAT) family ABC
15 putative glutamine ABC transporter, permease
26 ABC transporter membrane spanning permease
27 amino acid ABC transporter, amino acid-binding
17 COG0834: ABC-type amino acid transport/signal
29 COG0834: ABC-type amino acid transport/signal
30 GlnP
31 glutamine ABC transporter permease and substrate
28 amino acid ABC transporter, amino
55 Amino acid ABC transporter, permease protein
35 probable amino acid ABC transporter
13 Amino acid ABC transporter binding protein and
44 Pelodictyon luteolum DSM 273, complete genome
45 Amino acid ABC transporter, permease protein
78 Amino acid ABC transporter, permease protein
74 GlnQ

Figure 4. Phylogenetic tree for 100 sequences. Phylogenetic tree for 100 potential prokaryotic glutamate receptor channels as determined by presence of glutamate binding domain and transmembrane helices. (An electronic version of Figure 4 is included in supplementary materials to permit expansion for greater readability, Data S2.) The sequences are labeled with the definition line from the SDSC nonredundant protein database.
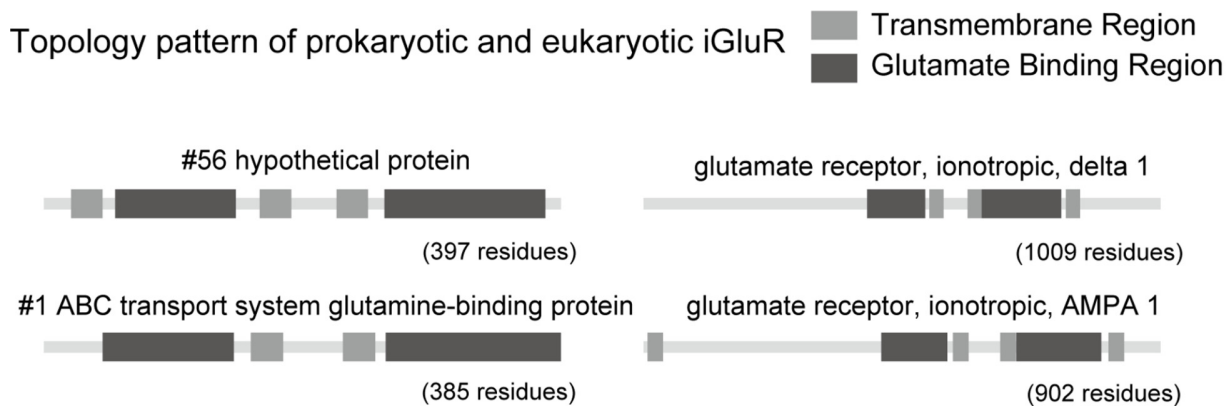doi:10.1371/journal.pone.0012827.g004

Figure 5. Topology pattern for Group 1 and Group 2. The eukaryotic counterpart of prokaryotic iGluR #56 is delta 1. The eukaryotic counterpart of prokaryotic iGluR #1 is AMPA 1. They are all with the structure of the S1 and S2 glutamate binding domains flanking two TM helices, in turn flanking a P-loop.
doi:10.1371/journal.pone.0012827.g005

horizontal gene transfer (HGT) between the ancestors of some proteobacteria and some cyanobacteria.

Comparison with eukaryotic glutamate receptor channels

Although iGluR research started with higher eukaryotic genomes, we still want to know if we can find all eukaryotic iGluR's by Group 1 and Group 2 sequences. First, we build a human iGluR list as a comparison by keyword search (Table 4).

Then, we used each of the Group1 and Group 2 as probes to blast human genome (BLASTP) [11], and accepted all hit with an e-value lower than 10. From the result (Table 5), we found that we can retrieve more human iGluR's using Group 2 as a probe. This implies that Group 2 sequences are closer to eukaryotic homologues than Group 1 sequences.

We also tested the orthologous relationship between eukaryotic iGluR prokaryotic iGluR by the "reciprocal-best-hits" criterion (data not shown). Both Group1 and Group2 members are orthologous to eukaryotic iGluR. This suggests two possible hypotheses. The first one is that Group 2 is the descendant of Group1 and eukaryotic iGluR is descendant of Group 2, because Group 2 is closer to eukaryotic iGluR in the phylogenetic map (data not shown). The other hypothesis is that eukaryotic iGluR is descendant of Group 2 and Group 1 is the combination of Group 2 and prokaryotic potassium channels.

Discussion

Our results have implications for gene annotation, microbial communication and the evolution of cellular communication, and the origin and evolution of circadian rhythms.

Gene Annotation

The gene products we identified as being homologous to ionotropic glutamate receptors are largely annotated otherwise. In this paper, we did individualized analysis to identify these gene

Table 2. Gene list of Group 1 and Group 2.

	Group 1	Protein ID
23	Possible ligand gated channel (GIC family	NP_896860.1
25	extracellular solute-binding protein, family	ZP_00674117.1
33	extracellular solute-binding protein, family 3	YP_378562.1
37	Ionoropic glutamate receptor	YP_376778.1
40	extracellular solute-binding protein, family 3	ABB23418.1
41	extracellular solute-binding protein, family	ZP_00517290.1
43	conserved protein of unknown function_ putative	YP_339120.1
46	extracellular solute-binding protein, family 3	ZP_00660701.1
56	hypothetical protein	NP_441171.1
62	Possible ligand gated channel (GIC family)	NP_894348.1
65	extracellular solute-binding protein, family	ZP_00530895.1
94	K channel, pore region	ZP_00533070.1
	Group 2	
1	ABC transport system glutamine-binding protein	NP_486951.1
2	COG0834: ABC-type amino acid transport/signal	ZP_00157839.2
3	Q3MEH3) Ionotropic glutamate receptor precursor	ABA20613.1
4	COG0834: ABC-type amino acid transport/signal	ZP_00108493.1
5	glutamine ABC transporter, periplasmic	YP_168531.1
6	COG0834: ABC-type amino acid transport/signal	ZP_00053934.2
9	extracellular solute-binding protein, family 3	ZP_00622239.1
42	glutamate-gated potassium channel	YP_204476.1
50	hypothetical protein	YP_132561.1
63	extracellular solute-binding protein, family 3	ZP_00629025.1

doi:10.1371/journal.pone.0012827.t002

Figure 6. Evolutionary Domain Network of 100 sequences. IPR001638: Bacterial extracellular solute-binding protein, family 3. IPR015683: Glutamate receptor-related. IPR000515: Binding-protein-dependent transport systems inner membrane component. IPR010065: Amino acid ABC transporter, permease protein, 3-TM region, His/Glu/Gln/Arg/opine. IPR001320: Ionotropic glutamate receptor. IPR013099: Ion transport 2. IPR003091: Voltage-dependent potassium channel. IPR001991: Sodium:dicarboxylate symporter. IPR002197: Helix-turn-helix, Fis-type. IPR000005: Helix-turn-helix, AraC type. IPR000408: Regulator of chromosome condensation, RCC1. IPR002052: N-6 Adenine-specific DNA methylase. IPR001508: NMDA receptor. IPR001828: Extracellular ligand-binding receptor. IPR015590: Aldehyde dehydrogenase.
doi:10.1371/journal.pone.0012827.g006

Table 3. Conservation comparison of Group 1 and Group 2.

		S1	S2	channel
Group 1	identical	10/97	1/132	17/115
Group 1	Strongly conserved	10/97	15/132	25/115
Group 1	Weakly conserved	9/97	12/132	10/115
Group 2	Identical	12/93	10/129	2/120
Group 2	Strongly conserved	16/93	17/129	15/120
Group 2	Weakly conserved	6/93	11/129	11/120

doi:10.1371/journal.pone.0012827.t003

products as likely ionotropic glutamate receptors. The key addition to the previous annotation comes from analysis by functional domains and by how those domains fit into the overall topology of the protein, especially where they are relative to the transmembrane helices. Our group has developed a high-throughput computational environment for such scanning (MotifNetwork) [23], based on the functional domain definitions in the InterPro database. MotifNetwork is being enhanced to consider topology as well, so we anticipate that the procedures described in this paper will ultimately be completely automated.

Microbial Communication and the Evolution of Cellular Communication

In previous work our group used domain analysis to discover previously unknown prokaryotic members of the Ach Receptor Ion Channel family [12], a discovery which was later experimen-

Sequence alignment in S1 region

```
41 LDVLIGPISITTER--FQKVAFTQPYFNAQIGLLVS
25 LDIIGPISITSER--LEKVAFTQPYFYAKIGLLAS
56 LDILIGPISVTPERAAIEGITFTQPYFSSGIGLLIP
37 IDVLVGPISITPRRLAMPGVDFTQPYYLAKSGVLLP
23 IDLLVGPISVTPDRLNLPGVDFTQPYFIGKEGILLP
62 IDLAIGPISITPDRVARNGIEFTQPYFYAEEGVLVP
43 VDLLAGPISITSER--VEKFLFSQPYYQSSLTIASR
09 ADMAAANISITAGR--EAVMDFSQPIFESGLQIMLH
05 ADMAIANISITAAR--ETEMDFSQPIFESGLQILVP
03 ANAGIAAISITAER--QQQFDFSLPMFSGGLQILVR
02 ANAGIAAISITAER--QQQFDFSLPMFSGGLQILVR
01 ANAGIAAISITAER--QQQFDFSLPMFSGGLQILVR
04 VNLGIAAISITAER--EQNFDFSLPIFASGLQIMVR
06 ADLAISAISITAQR--EAQFDFSHPMYDSGLAILVQ
46 VDIAAAALTMTVER--EEQLDFSHPYFQSGLAIAVR
40 VDLAAAALTMTVER--EKQLDFSHPYFQSGLAIAVR
33 VDAAAAAITITKER--EYLLDFSAPYFHSGLAIAVK
65 LDAGVAAITVTAER--EETLDFSQPYYLSRFGIATL
50 INIGISCISITPER--EERLDFSHSFYETHLAIAVK
42 IDVGVSCISITPDR--EEHVDFSHSFYETHLAIAVK
94 IDIAISPLTVTASR--MKKFSFSQPYYITNLAFAMK
63 VDAAVTNLTITRER--AELIAFTQPWYDAGLRIMVP
                 *
```

Sequence alignment in S2 region

```
03 QKISVLEVPKIEQAYDALETKKAEAVVFDS
02 QKISVLEVPKIEQAYDALETKKAEAVVFDS
01 QKISVLEVPKIEQAYDALETKKAEAVVFDS
04 HHISVLEVPKIEEAYKALQTKKADAVVFDA
06 LRVDTLVFPTVAEAIASTESGKTSAVVYDS
09 REIDYVAFPGLDKMLEDFEDGDTRIVVFDA
05 RDIDFAAFSDLQEMITAFERGALDAIVFDA
63 MGIATRSYVNIDASVVALREGRLDAVVGDA
37 GSMRVLKTENLTSGIELVLSEQAEAVIFDR
23 RNMRIVPAKTLTAAIDHVLMNRAEAVIFDR
62 YGARPFPVPTLKEAIHLIKRNKVSGVISDT
41 YQSRVLPSPNLEQAIERLKSGQAEGVMFDV
25 YEARINQTETLVDAINLLKLNEVDGVIFDV
56 YQADVRETNNLTAAITLLQKKQVEAVMFDR
50 HSIVHKAYNTVEDMLVALEKGEIEAVIADD
42 HSIAHDTYKDMDALLLALEKGQLDAVVADD
46 LAIRGQNFPTVNDALHLLEQGKLDAVVYDE
40 LAIKGQYFPTVKEAMQALEKGQIDAVVYDE
33 STMKSKHFLTIQDALHALENDQIDAVVFDE
65 EYIAFQHFATIQDALQALQERRVEAVLYDE
94 FKIKFVDFSTVEDGLVAVQKDEIAAFVYDH
43 HQGQKVLVNNLSQAMERLKDKSVDGVVFDR
                              *
```

Figure 7. The sequence alignment in S1 and S2 region of Group1 and Group 2 proteins.
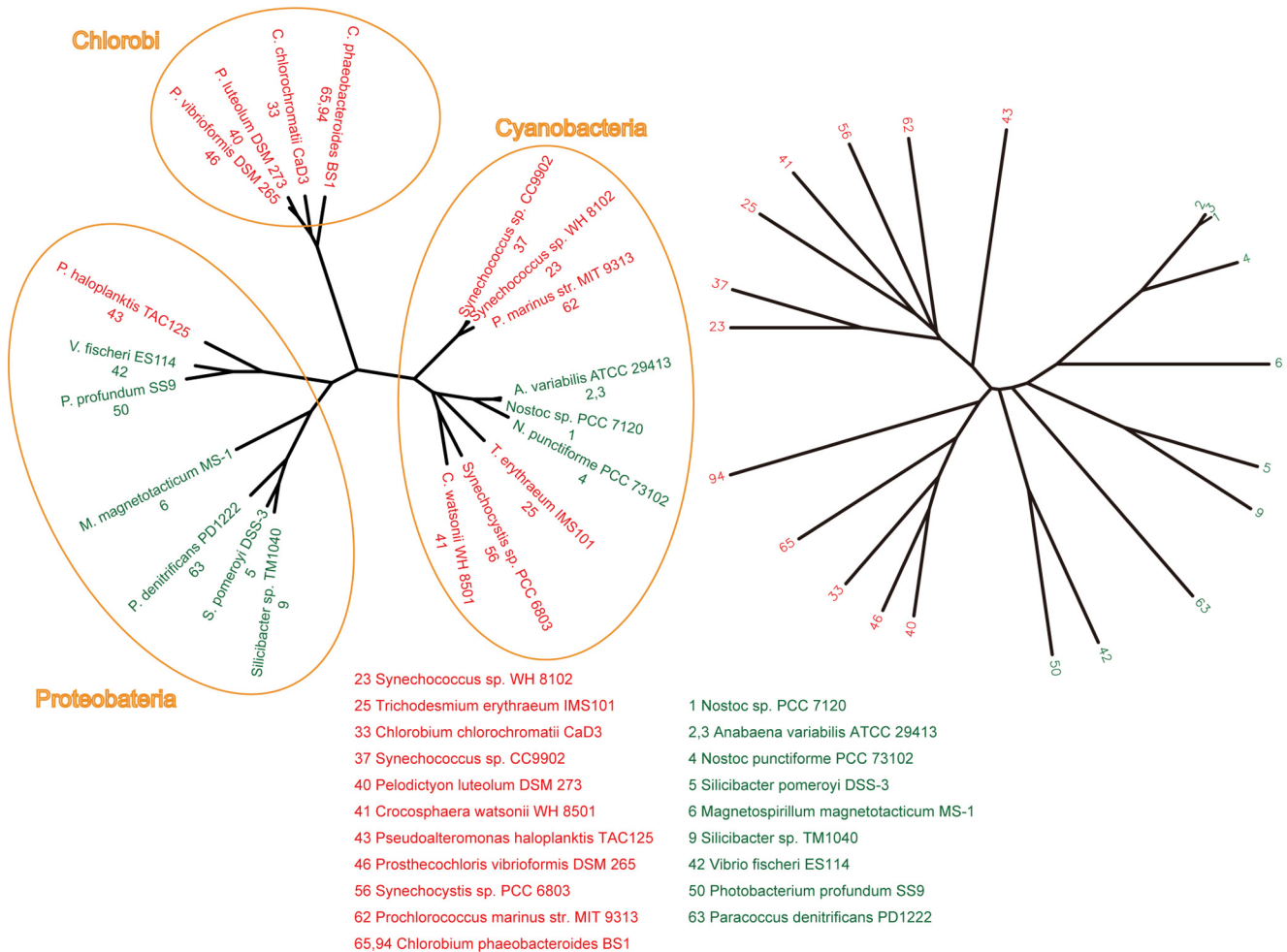doi:10.1371/journal.pone.0012827.g007

Figure 8. Phylogenetic trees of 16s rRNA genes and Group 1/Group 2 genes. Left hand side is the 16sRNA tree for the species that contain Group 1 and Group 2 prokaryotic glutamate receptor channels. Right hand side is the tree for the Group 1 and Group 2 proteins. The fact that the clustering patterns are different for the two trees indicates horizontal gene transfer of glutamate receptor channels among the bacteria. In particular, it seems there must have been a minimum of two transfers, one from cyanobacteria to proteobacteria, and one from proteobacteria to cyanobacteria.
doi:10.1371/journal.pone.0012827.g008

Table 4. Human iGluR's.

AMPA	AMPA 1	NP_000818.1	906 aa
AMPA	AMPA 2 isoform 1	NP_000817.2	883 aa
AMPA	AMPA 2 isoform 2	NP_001077088.1	883 aa
AMPA	AMPA 2 isoform 3	NP_001077089.1	836 aa
AMPA	glutamate receptor 3 isoform flip	NP_015564.4	894 aa
AMPA	glutamate receptor 3 isoform flop	NP_000819.3	894 aa
AMPA	AMPA 4 isoform 1	NP_000820.3	902 aa
AMPA	AMPA 4 isoform 2	NP_001070711.2	884 aa
Kainate	kainate 1 isoform 1	NP_000821.1	918 aa
Kainate	kainate 1 isoform 2	NP_783300.1	905 aa
Kainate	kainate 2 isoform 1	NP_068775.1	908 aa
Kainate	kainate 2 isoform 2	NP_786944.1	869 aa
Kainate	kainite 3	NP_000822.2	919 aa
Kainate	glutamate receptor KA1	NP_055434.2	956 aa
Kainate	glutamate receptor KA2	NP_002079.3	980 aa
NMDA	NMDA receptor 1 isoform NR1-1	NP_000823.4	885 aa
NMDA	NMDA receptor 1 isoform NR1-2	NP_067544.1	901 aa
NMDA	NMDA receptor 1 isoform NR1-3	NP_015566.1	938 aa
NMDA	N-methyl-D-aspartate receptor subunit 2A	NP_000824.1	1464 aa
NMDA	N-methyl-D-aspartate receptor subunit 2D	NP_000825.2	1336 aa
NMDA	N-methyl-D-aspartate receptor subunit 2C	NP_000826.2	1233 aa
NMDA	N-methyl-D-aspartate receptor subunit 2B	NP_000827.2	1484 aa
NMDA	N-methyl-D-aspartate 3A	NP_597702.1	1115 aa
NMDA	N-methyl-D-aspartate 3B	NP_619635.1	1043 aa
Delta	delta 1	NP_060021.1	1009 aa
Delta	delta 2	NP_001501.2	1007 aa

doi:10.1371/journal.pone.0012827.t004

Table 5. Reverse BLAST result against human genome using Group 1 and Group2 as a probe.

Group 1		Group 2	
protein No.	ratio	protein No.	ratio
23	13/26	1	26/26
25	26/26	2	26/26
33	26/26	3	26/26
37	13/26	4	26/26
40	26/26	5	26/26
41	24/26	6	25/26
43	19/26	9	26/26
46	26/26	42	25/26
56	16/26	50	20/26
62	14/26	63	18/26
65	26/26		
94	26/26		

doi:10.1371/journal.pone.0012827.t005

tally confirmed [13]. In this paper we extend the work to another major group of ligand-gated channels, the glutamate receptor channel family. These two discoveries together contribute to larger questions. What is the evolutionary origin of the electrochemical signaling mechanisms utilized in neuronal, neuromuscular, and neuroendocrine systems? To what extent do contemporary prokaryotes use these mechanisms to communicate? It should be noted that the patterns of occurrence of the two families of ligand-gated channels are very different. The prokaryotic Ach receptor channels are distributed across widely varying types of prokaryotes, both bacteria and archaea. By contrast, we found glutamate receptor channels only in bacteria, and clustered in particular bacterial subgroups. Because the sequence coverage of microbial genomes is still so sparse relative to the full range of microbial diversity, it is not possible to assess the full significance of this contrast. Based on our analysis of the existing data, it appears that horizontal transfer was the major mechanism for disseminating the prokaryotic members of the Ach receptor channel family. The members of the glutamate receptor channel family show evidence of at least two incidents of horizontal transfer (see Figure 8) but otherwise disseminate and variegate by descent. Based on the evolutionary domain network of the prokaryotic channels, it appears that domain reorganization was a significant factor in their evolution.

Origin and Evolution of Circadian Rhythms

We note three facts:

1) Among all prokaryotes, cyanobacteria have been shown to exhibit circadian rhythms [24].

2) In this paper, we find that among prokaryotes, ionotropic glutamate receptor channels are disproportionately present in cyanobacteria.

3) In animal brain slice preparations, glutamate resets circadian rhythms in a manner similar to light [25].

From this combination of facts, we are moved to suggest that glutamate signaling may provide a link connecting the circadian regulation of animals and cyanobacteria. This suggestion needs to be tested by further work.

Supporting Information

Data S1 Detailed information for 100 sequences included in this analysis.
Found at: doi:10.1371/journal.pone.0012827.s001 (0.10 MB PDF)

Data S2 Phylogenetic tree for 100 sequences included in this analysis.
Found at: doi:10.1371/journal.pone.0012827.s002 (0.78 MB PDF)

Data S3 Topology patterns for 100 sequences included in this analysis.
Found at: doi:10.1371/journal.pone.0012827.s003 (9.04 MB PDF)

Data S4 Sequence alignments of S1, S2 and channel regions in Group 1 and Group 2.
Found at: doi:10.1371/journal.pone.0012827.s004 (2.00 MB PDF)

Data S5 Phylogenetic trees of S1, S2 and channel regions in Group 1 and Group 2.

Acknowledgments

Author Contributions

Conceived and designed the experiments: MFG EJ. Performed the experiments: MFG GR. Analyzed the data: MFG GR JT. Contributed reagents/materials/analysis tools: GR JT. Wrote the paper: MFG EJ.

References

1. Stevens TJ, Arkin IT (2000) Do more complex organisms have a greater proportion of membrane proteins in their genomes? Proteins 39: 417–420.
2. Gouaux E, Mackinnon R (2005) Principles of selective ion transport in channels and pumps. Science 310: 1461–1465.
3. Pohorille A, Schweighofer K, Wilson MA (2005) The origin and early evolution of membrane channels. Astrobiology 5: 1–17.
4. Ashcroft FM (2006) From molecule to malady. Nature 440: 440–447.
5. Chen GQ, Cui C, Mayer ML, Gouaux E (1999) Functional characterization of a potassium-selective prokaryotic glutamate receptor. Nature 402: 817–821.
6. Mayer ML (2006) Glutamate receptors at atomic resolution. Nature 440: 456–462.
7. Booth IR, Edwards MD, Miller S (2003) Bacterial ion channels. Biochemistry 42: 10045–10053.
8. Kuner T, Seeburg PH, Guy HR (2003) A common architecture for K+ channels and ionotropic glutamate receptors? Trends Neurosci 26: 27–32.
9. Kung C, Blount P (2004) Channels in microbes: so many holes to fill. Mol Microbiol 53: 373–380.
10. Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, et al. (1998) The structure of the potassium channel: molecular basis of K+ conduction and selectivity. Science 280: 69–77.
11. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.
12. Tasneem A, Iyer LM, Jakobsson E, Aravind L (2005) Identification of the prokaryotic ligand-gated ion channels and their implications for the mechanisms and origins of animal Cys-loop ion channels. Genome Biol 6: R4.
13. Bocquet N, Prado de Carvalho L, Cartaud J, Neyton J, Le Poupon C, et al. (2007) A prokaryotic proton-gated ion channel from the nicotinic acetylcholine receptor family. Nature 445: 116–119.
14. Hilf RJ, Dutzler R (2008) X-ray structure of a prokaryotic pentameric ligand-gated ion channel. Nature 452: 375–379.
15. Saier MH, Jr. (2000) A functional-phylogenetic classification system for transmembrane solute transporters. Microbiol Mol Biol Rev 64: 354–411.
16. Mayer ML, Olson R, Gouaux E (2001) Mechanisms for ligand binding to GluR0 ion channels: crystal structures of the glutamate and serine complexes and a closed apo state. J Mol Biol 311: 815–836.
17. Moller S, Croning MD, Apweiler R (2001) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17: 646–653.
18. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
19. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package). 3.5c ed.
20. Hu Z, Frith M, Niu T, Weng Z (2003) SeqVISTA: a graphical tool for sequence feature visualization and comparison. BMC Bioinformatics 4: 1.
21. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783–795.
22. Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847–848.
23. Tilson JL, Rendon G, Ger M-F, Jakobsson E (2007) MotifNetwork: A Grid-enabled Workflow for High-throughput Domain Analysis of Biological Sequences. Bioinformatics and Bioengineering, 2007 BIBE 2007 Proceedings of the 7th IEEE International Conference. pp 620–627.
24. Mackey SR, Golden SS (2007) Winding up the cyanobacterial circadian clock. Trends Microbiol 15: 381–388.
25. Kalsbeek A, Kreier F, Fliers E, Sauerwein HP, Romijn JA, et al. (2007) Minireview: Circadian control of metabolism by the suprachiasmatic nuclei. Endocrinology 148: 5635–5639.