

RESEARCH

Open Access



Amino acid sequence associated with bacteriophage recombination site helps to reveal genes potentially acquired through horizontal gene transfer

Maria A. Daugavet^{1*} , Sergey V. Shabelnikov¹ and Olga I. Podgornaya^{1,2,3}

From Bioinformatics: from Algorithms to Applications 2019 (BIATA 2019) conference
Saint Petersburg, Russia. 20-22 June 2019

* Correspondence: ka6tanka@yandex.ru

¹Institute of Cytology, St. Petersburg, Russia 194064
Full list of author information is available at the end of the article

Abstract

Background: Horizontal gene transfer, i.e. the acquisition of genetic material from nonparent organism, is considered an important force driving species evolution. Many cases of horizontal gene transfer from prokaryotes to eukaryotes have been registered, but no transfer mechanism has been deciphered so far, although viruses were proposed as possible vectors in several studies. In agreement with this idea, in our previous study we discovered that in two eukaryotic proteins bacteriophage recombination site (AttP) was adjacent to the regions originating via horizontal gene transfer. In one of those cases AttP site was present inside the introns of cysteine-rich repeats. In the present study we aimed to apply computational tools for finding multiple horizontal gene transfer events in large genome databases. For that purpose we used a sequence of cysteine-rich repeats to identify genes potentially acquired through horizontal transfer.

Results: HMMER remote similarity search significantly detected 382 proteins containing cysteine-rich repeats. All of them, except 8 sequences, belong to eukaryotes. In 124 proteins the presence of conserved structural domains was predicted. In spite of the fact that cysteine-rich repeats are found almost exclusively in eukaryotic proteins, many predicted domains are most common for prokaryotes or bacteriophages. Ninety-eight proteins out of 124 contain typical prokaryotic domains. In those cases proteins were considered as potentially originating via horizontal transfer. In addition, HHblits search revealed that two domains of the same fungal protein, Glycoside hydrolase and Peptidase M15, have high similarity with proteins of two different prokaryotic species, hinting at independent horizontal gene transfer events.

(Continued on next page)



(Continued from previous page)

Conclusions: Cysteine-rich repeats in eukaryotic proteins are usually accompanied by conserved domains typical for prokaryotes or bacteriophages. These proteins, containing both cysteine-rich repeats, and characteristic prokaryotic domains, might represent multiple independent horizontal gene transfer events from prokaryotes to eukaryotes. We believe that the presence of bacteriophage recombination site inside cysteine-rich repeat coding sequence may facilitate horizontal genes transfer. Thus computational approach, described in the present study, can help finding multiple sequences originated from horizontal transfer in eukaryotic genomes.

Keywords: Bacteriophages, Recombination site, Protein domains, Horizontal gene transfer, Cysteine-rich repeats

Background

As a general rule, genetic material is inherited by an offspring from its parent. This type of gene transfer is called vertical. Another way of gene flow is horizontal gene transfer (HGT), which means the acquisition of DNA from non-related species [1]. This phenomenon is widely accepted for prokaryotes [2, 3]. Mechanisms of HGT in prokaryotes are well studied and include transformation, conjugation and transduction [4]. Multiple cases of gene transfer from prokaryotes to eukaryotes have also been registered [5, 6]. For instance, one of the mechanisms of DNA transfer from *Agrobacterium* species to its plant host involves type IV secretion systems (T4SSs) of bacteria [7]. It was also shown that bacteria *Escherichia coli* [8] can deliver DNA via conjugation-like mechanism to cultured eukaryotic cells under artificial conditions [9, 10]. Nevertheless, no such mechanisms as are common between prokaryotes have been discovered so far for metazoans in nature. Yet the importance of HGT for latter may be quite significant [7, 11]. For example, the acquisition of the lysyl oxidase enzyme, one of the metazoan synapomorphy, might have involved a prokaryote source [12]. HGT is also suspected to contribute to the fast Cambrian radiation of Metazoa [13, 14]. Virus involvement as a carrier of foreign DNA [15] was proposed for many cases of horizontal transfer of transposons [16] and protein coding sequences [17–19]. Phage lambda is believed to be involved in DNA exchange between bacteria and human somatic cells [20]. *Escherichia coli* PK1A2 bacteriophage was shown to penetrate into eukaryotic neuroblastoma cells under experimental conditions [21], albeit nuclear delivery of DNA was not detected in that study. For some viruses, the presence of the nuclear localization signals was shown in their terminal proteins. Those nuclear localization signals proved to be functional and to facilitate gene delivery into the eukaryotic nucleus [22].

Cases of HGT from prokaryotes to eukaryotes are common among Fungi [23–25] and unicellular organisms [26–29]. They are less frequent among metazoans, however, with some groups more prone to HGT from prokaryotes than others. Increased HGT susceptibility may be due to asexual reproduction [30] and/or to the contact of gametes with the environment. Multiple cases of HGT were reported, for example, in nematodes [31–33], rotifers [34–36] and cnidarians [37]. Tunicates, a basal chordate group, are exceptional in their use of various ways of asexual reproduction [38, 39]. Previous studies revealed two proposed cases of HGT from prokaryotes to tunicates [40, 41]. The first one is the cellulose-synthase gene of ascidians, which was gained from bacterial donor *Streptomyces* sp. [40]. The second case involves a possibly chimeric protein

rusticalin from ascidian *Styela rustica*, in which the coding sequence of the C-terminal domain might have been inherited from bacteriophage A500 [41].

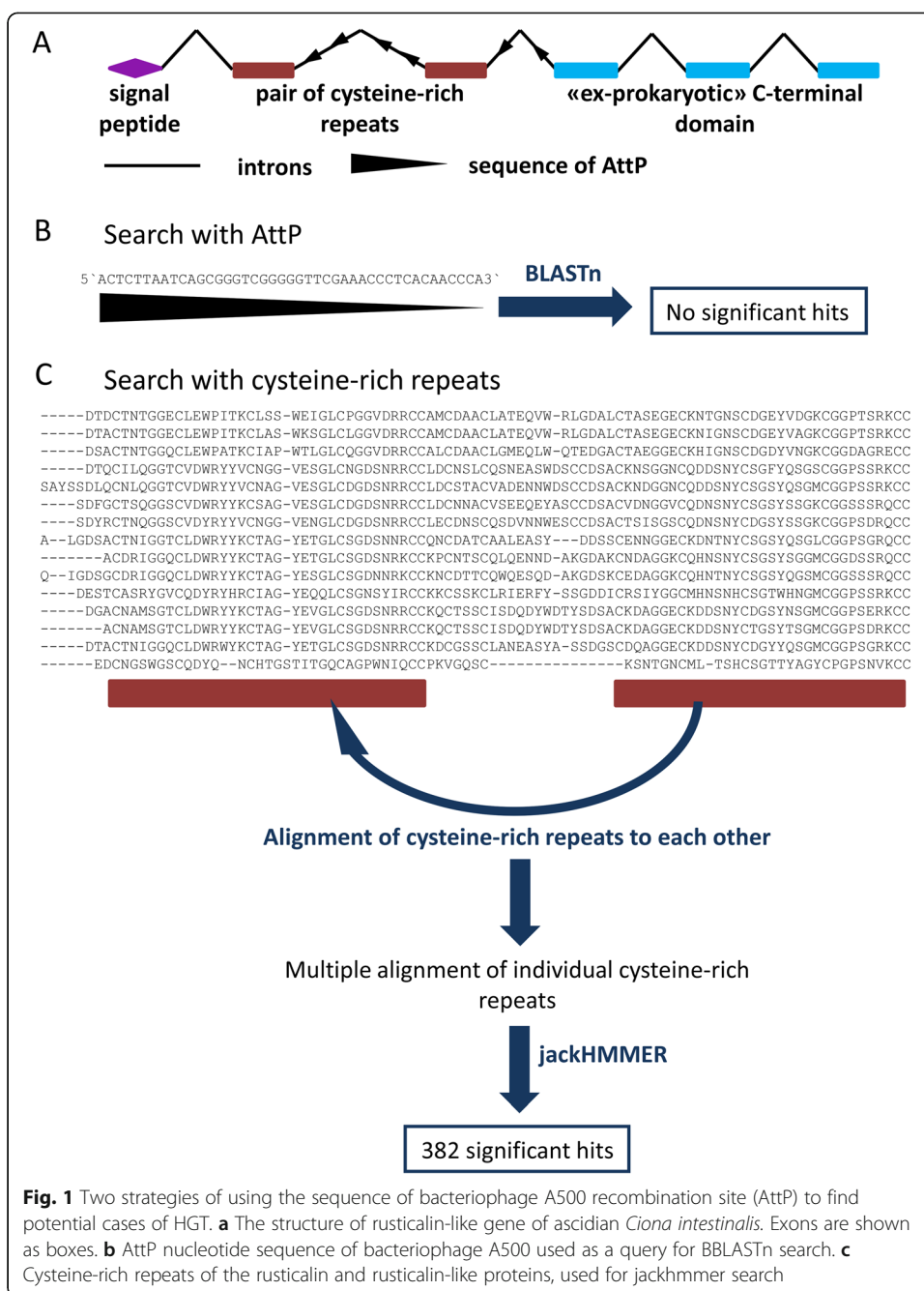
Rusticalin was described as a hyalinocytes-specific protein of *Styela rustica*. The only discernible homologues of rusticalin were found in basal chordates, corals, and placozoans. According to the predicted features, based on the sequence analysis, rusticalin should consist of two distinct regions, the N-terminal domain and the C-terminal domain. The N-terminal domain comprises two cysteine-rich repeats and shows remote similarity to the tick carboxypeptidase inhibitor and also to β -defensin anti-bacterial peptides. The C-terminal domain, on the other hand, shares significant sequence similarity with bacterial MD peptidases and bacteriophage A500 L-alanyl-D-glutamate peptidase. Thus, the N-terminal domain of rusticalin comprises two cysteine-rich repeats of supposedly eukaryotic origin, and C-terminal domain potentially has a prokaryotic origin [41]. The coding region of the N-terminal domain contains introns with possible bacteriophage recombination sites (AttP) hidden inside, which means that C-terminal domain is adjacent to a possible AttP(s). Both sequence similarity and the presence of a putative bacteriophage recombination site support the hypothesis of the C-terminal domain coming from bacteriophage genome.

In another example of HGT in ascidians, the coding region of the cellulose-synthase catalytic domain also neighbored a sequence similar to bacteriophage recombination site. Based on these results, we proposed that a cellulose synthase catalytic subunit was acquired through the similar mechanism, involving bacteriophage as a vector. Thus, our previous work suggested a possible HGT mechanism involving bacteriophage insertion in at least two cases of transfer [41]. In the present study we aimed to find additional cases of HGT using the sequence of bacteriophage recombination site. Since the nucleotide sequence of AttP is too short, we used instead the amino acid sequence (cysteine-rich repeats) harboring AttP as a possible marker of transfer events inside eukaryotic chromosomes. About a hundred of proteins that possibly originated through HGT were found in this way.

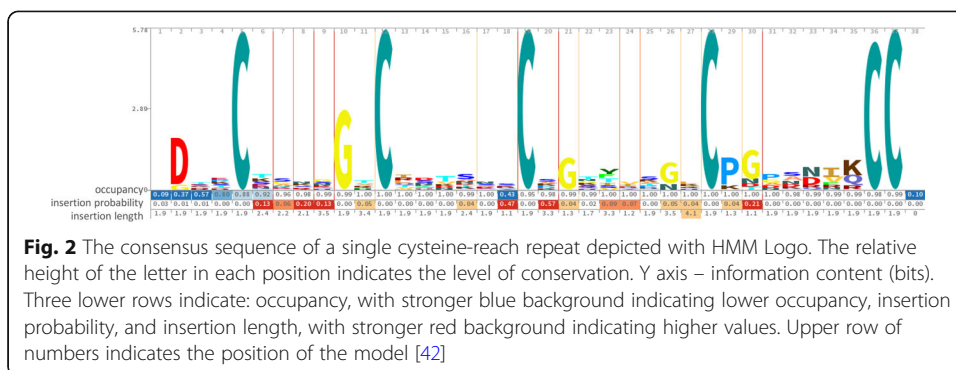
Results

Proteins containing cysteine-rich repeats

Our approach at finding bacteriophage recombination sites using BLASTn search through all available eukaryotic genomes returned no significant hits (Fig. 1b). Since our previous results indicated that AttP sites lie inside the introns of rusticalin cysteine-rich repeats (Fig. 1a), we switched to using a larger amino acid sequence of cysteine-rich repeat itself for a remote similarity search. For that purpose, we split each cysteine-rich repeat, present in rusticalin and rusticalin-like proteins, into individual cysteine-rich modules. Those modules were aligned to each other in order to obtain a multiple sequence alignment (Fig. 1c). This alignment of individual cysteine-rich modules was used in a remote similarity search by JackHMMER in UniProtKB database. Three iterations of jackhmmmer gave a maximum number of hits with a small number of hits losses. It resulted in 382 significant matches with protein sequences (Supplementary material, Table T1) with E-value ranging from $1.4e-207$ to 0.0098.



Resulting protein dataset was used for future analysis. For all proteins containing cysteine-rich repeats an HMM Logo of an individual repeat was constructed (Fig. 2), showing the level of conservation for each amino acid position. Cysteine pattern appeared to be absolutely conserved. Among the other conservative amino acids there are amino acids with small neutral side chains like glycine (positions 10, 21, 26 and 30) and proline (positions 29 and 31). This sequence was usually present in multiple copies. The taxonomic distribution of the proteins containing one or more cysteine-rich repeats is given in Fig. 3. Such proteins are almost exclusively eukaryotic (374 out of 382): majority of the proteins (301 out of 382) belong to Fungi and only 69 to Metazoa

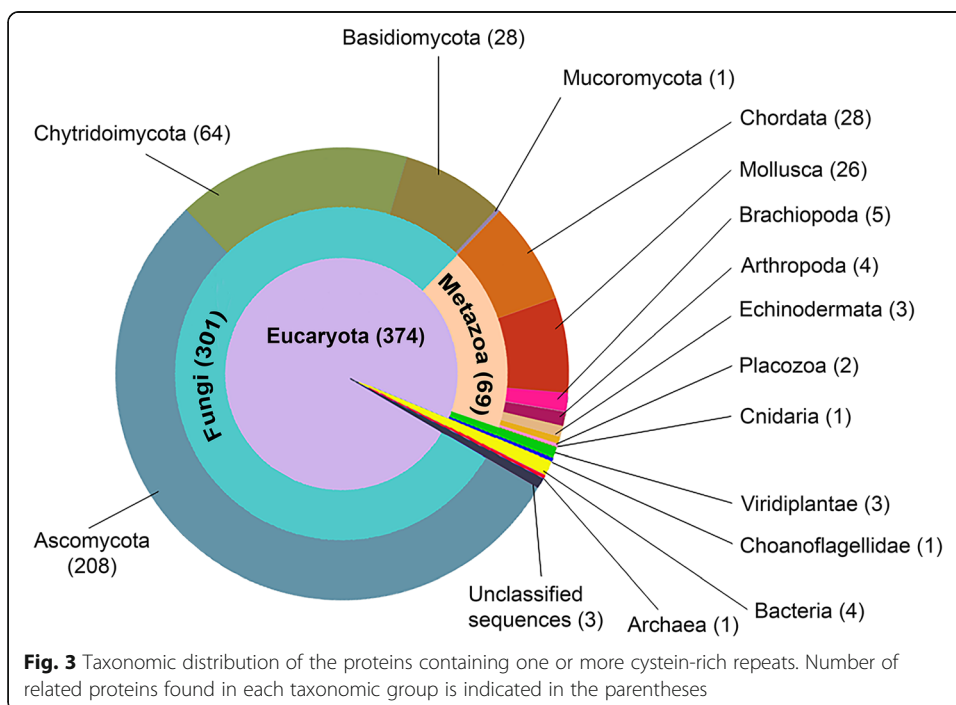


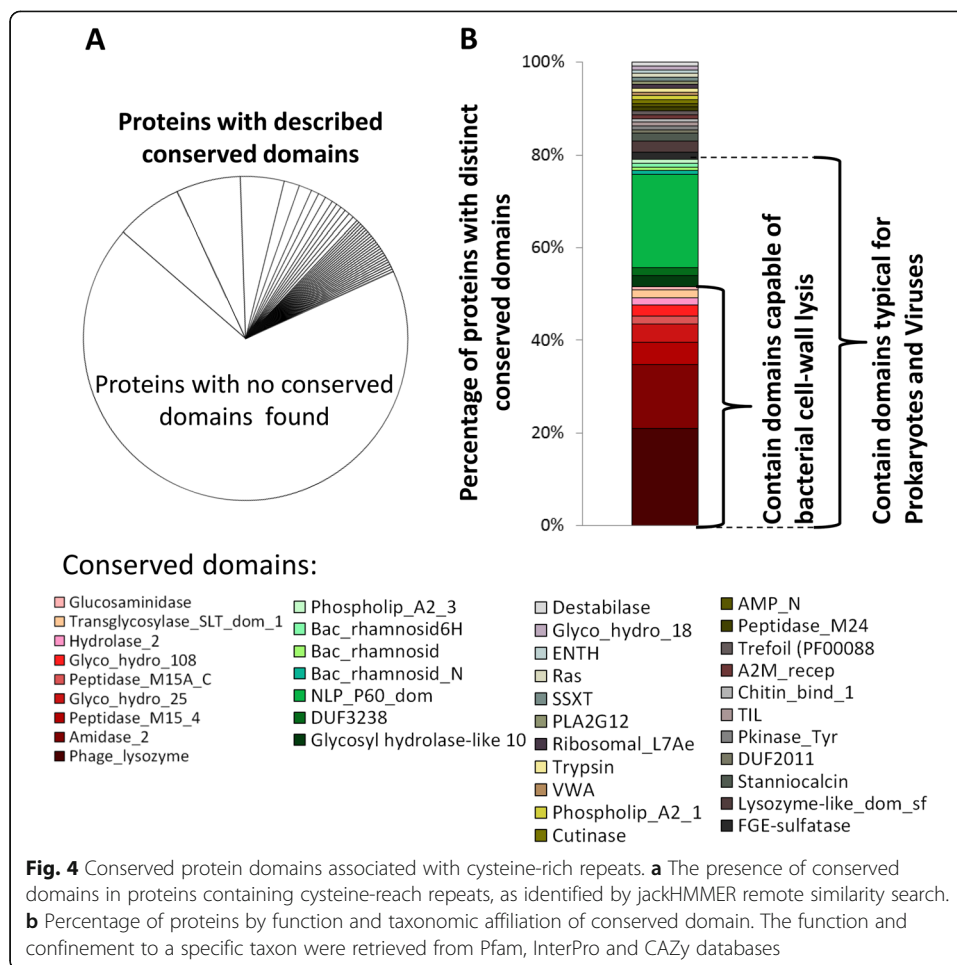
(Fig. 3). The distribution of hits among metazoan taxa is patchy. We found a wide variety of phyla but low abundance of the proteins in each phylum.

Conserved domains associated with cysteine-rich repeats

Cysteine-rich repeats are usually found as parts of larger proteins. In our study 124 proteins had such repeats present together with other annotated conserved domains (Fig. 4a). These proteins formed a restricted dataset which was used in further analysis. In the other 258 matches cysteine-rich repeats were present, but no associated annotated domains were found nearby. We believe these proteins must be analyzed separately and deserve a dedicated study.

Among 124 proteins containing cysteine-rich repeats and predicted conserved domains, in 20% (26 proteins) such repeats were associated with phage-lysozyme (PF00959), in 14% (17 proteins) with zinc amidase (PF01510) and in 5% (6 proteins) with Peptidase M15 (PF13539). In a few other proteins the cysteine-rich repeats were





associated with other domains (Fig. 4b). For each conserved domain we screened its global species distribution with automatic Pfam description [43]. We identified a total of 16 different domains that can be classified as typical for prokaryotes or bacteriophages and they are present in 79% (98 out of 124) of proteins with characterized conserved domains. Remaining 21% of proteins containing annotated domains had no bias towards prokaryotes in their taxonomic distributions. The search for physiological functions of the possible prokaryotic or viral domains found revealed that nine are bacterial cell-wall hydrolyzing enzymes and they are present in 51% of proteins (63 out of 124) (Fig. 4b). Other seven domains are either not involved in cell-wall destruction, or their functions are unknown. Nevertheless all proteins containing domains typical for prokaryotes or bacteriophages may be considered as candidates for being originated through HGT process.

New case of HGT

In a restricted protein dataset we identified several HGT candidates based on the description of conserved domains. In order to identify the potential records of multiple transfer events we chose proteins with more than one predicted conserved domains. Two proteins from the fungus *Neocallimastix californiae* (UniProt ID A0A1Y2AHN7

hydrolase and Peptidase M15 domains in the coding sequences of *N. californiae* protein might have originated from a prokaryotic ancestor.

The same logic is applicable for the protein A0A1Y2AHN7 containing Glucosaminidase (PF01832) coupled with Endopeptidase (PF000877). Those two domains might have come from two independent events of HGT, but they also could have been transferred as a single DNA fragment from a single prokaryotic donor organism.

Discussion

The only mechanism of prokaryote-to-eukaryote DNA transfer established so far involves a bacterial pathogen donor and a plant host as an acceptor. In this case DNA is delivered into the nucleus and integrated into the chromosome by the host DNA repair machinery, summarized in [7]. Another model of HGT involving bacteriophage as a vector of gene transfer from prokaryote donor to eukaryote acceptor was proposed earlier [15, 21, 22, 44, 45]. In particular it has been described that horizontally acquired genes were associated with prophage regions in the donor *Wolbachia* genome [46]. The results of our previous study agree with these findings. The presence of the bacteriophage recombination sites (AttP) next to horizontally transferred genes in eukaryotic genome as well as in bacterial donor genome supports this hypothesis. In the present study we demonstrated that a search for bacteriophage recombination site in eukaryotic genomes can reveal new cases of HGT. Although, since the nucleotide sequence of AttP is too short to get significant hits, we switched to a larger amino acid sequence of cysteine-rich repeats harboring AttP inside its introns. Cysteine-rich repeats happened to be conservative across multiple fungal and metazoan proteins. The similarity between cysteine-rich repeats and β -defensins may suggest that both are involved in immune reaction [47–49]. The distribution of the proteins with cysteine-rich repeats among Metazoa is patchy and not concentrated in any particular phylogenetic group. This disjointed distribution was previously described as a hallmark of horizontally acquired genes [27, 28, 45] and was also used as an instrument to find transposon horizontal transfers [50].

Here we further analyzed domain architecture of the proteins containing cysteine-rich repeats. In 124 cases conserved domains were predicted to accompany cysteine-rich repeats, while for 258 remaining proteins prediction was unsuccessful. This may be due to the low conservation of amino acid sequences between different taxa [51, 52]. It is also possible that other prediction instruments [53] such as Motif Scan (https://myhits.isb-sib.ch/cgi-bin/motif_scan) or MOTIF search (<https://www.genome.jp/tools/motif/>) would be more sensitive than HMMER annotation we used here. Some of the proteins with no predicted domain architecture belong, nevertheless, to the species with previously described multiple cases of HGT. Examples include the fungi *Pochonia chlamydosporia* probably harboring 100 kb region of foreign DNA [54], *Fusarium oxysporum* [55] and *Metarhizium majus* [56]. Cysteine-rich repeats found in these species proteins might provide a bacteriophage dependent mechanism for such HGT events.

In a restricted dataset of the 124 proteins containing predicted conserved domains, we screened the taxonomic distribution of each domain using Pfam and InterPro databases. Even though cysteine-rich repeats themselves are found almost exclusively in eukaryotic proteins, their associated domains, which we were able to identify, were often typical for prokaryotes or bacteriophages. Such associations constituted 79% of

the hits in our restricted dataset. Based on this high incidence, we hypothesize that such domains are originated through HGT. Moreover, the cutinase domain, which is uniformly present in prokaryotes and eukaryotes, was likely transferred laterally from Bacteria to Fungi [27]. Thus, we may even underestimate the proportion of HGT domains in our dataset.

Some domains associated with cysteine-rich repeats were described earlier as HGT participants. For example, phage lysozyme was found to be horizontally transferred in bivalve mollusks genome [57] and Glycoside hydrolase domain was probably inserted independently into multiple genomes: in Bacteriophages, Archaea and in three clades of Eukarya [44]. We found Peptidase M15 domain in Fungi and Metazoa (*Trichoplax adhaerens*) and Amidase_2 domain in tree lineages of Metazoa (Chordata, Molluska and Arthropoda) (Supplementary Table T1). In those cases we can also hypothesize independent transfer events.

Two of the proteins in our dataset happened to contain two different predicted domains with suggested “ex-prokaryotic” origins. Such unusual domain architecture leads us to assumption of chimeric origin of these proteins, where the coding sequences of individual domains could have been inherited from prokaryote donors. In many cases of HGT among bacteria, it is the protein domains rather than whole genes, considered as units of transfer [58, 59]. According to our results, two domains of the fungal protein A0A1Y2FMX2 show significant similarity (E-value: 2.0E-28 and 2.6E-33) to bacterial sequences. At the same time no other closely related proteins were found among other taxa. Glycoside hydrolase domain has the putative homolog sequence in the genome of bacteria *Lachnospirillum* sp., while the most similar sequence to Peptidase M15 domain was found in the genome of bacteria *Bacteroides clarus*. Each of these domains in a protein sequence was accompanied by a pair of cysteine-rich repeat. It is worth mentioning that the genus *Lachnospirillum* belongs to the phylum Firmicutes, while the genus *Bacteroides* belongs to the phylum Bacteroidetes. Thus, two probable bacterial donors occupy very distant phylogenetic positions [60]. This fact suggests that there might have been two independent HGT events which created a protein with two “ex-prokaryotic” domains.

Both Glycoside hydrolase and Peptidase M15 are enzymes capable of bacterial cell wall lysis [61–64]. We also found a bias towards bacterial cell-wall destruction among the functions of the other predicted domains. Cysteine-rich repeats might serve as antimicrobial peptides penetrating bacterial cell wall in conjunction with lytic enzymes. Such conjunction may even give the organism an immediate selective advantage in antibacterial defense.

Many other described cases of HGT involve a variety of enzymes [59, 65, 66] covering a broad range of metabolic functions [6], whereas proteins predicted in our study as HGT cases are largely supposed to be cell-wall lytic enzymes. Taking into account that bacteriophages use cell-wall lytic enzymes during the replication cycle [67, 68], we hypothesize bacteriophage involvement as a vector of transfer. In this case a new foreign protein would be carried not only as a sequence residing in viral genome but also might serve as a functional enzyme for a prolonged period of time before its horizontal transfer into a eukaryotic cell. Numerous cell-wall destruction enzymes, found in this study, might serve as indirect evidence that bacteriophage was a transition step in gene transfer. We previously hypothesized that the sequence of bacteriophage recombination

site (AttP), located inside cysteine-rich repeats, can facilitate a type of HGT which involves bacteriophage as a vector of gene flow. Among the proteins described in the present study, some contain domains typical for bacteriophages, but no direct homology was found. This is probably due to the fast evolutions of viral genomes [69, 70] which can mask the similarity of related proteins [71]. Nevertheless, cysteine-rich repeats can serve as an instrument to find new cases of prokaryote to eukaryote HGT. We also demonstrated that a split of amino acid sequence according to the predicted domain borders may help to infer the ancestry for each domain separately and detect HGT cases.

Conclusions

Cysteine-rich repeats in eukaryotic proteins are usually accompanied by conserved domains typical for prokaryotes or bacteriophages. Those chimeric proteins probably represent multiple independent HGT events from prokaryotes to eukaryotes. The explanation of this phenomenon may lie in the presence of bacteriophage recombination site, which potentially facilitates HGT, inside the coding sequence of the cysteine-rich repeats.

Methods

Constructing the dataset of the proteins containing cysteine-rich repeats

In order to find HGT candidates, we searched all eukaryotic genomes present in the nr, est and TSA GenBank databases for the nucleotide sequence of bacteriophage AttP using BLASTn. Since it provided no significant hits, we processed amino acid sequences of cysteine-rich repeats for subsequent remote similarity search instead. Cysteine-rich repeats are present in rusticalin and rusticalin-like proteins as a pair [41]. The borders of repeats in this amino acid sequence were predicted using REPRO [72]. According to those borders, each member of repeat pair was split into two individual cysteine-rich modules. At the next step we aligned all modules of all rusticalin-like proteins with MUSCLE 3.8.31 [73]. Multiple sequence alignment of individual cysteine-rich modules was subjected to remote similarity searches using online version of HMMER 3.1b2 jackhmmer [74] in UniProtKB v.2017_08 protein database. The resulting list of hits became a raw dataset of the proteins containing cysteine-rich repeats.

Conserved domains analysis

Taxonomic distribution of proteins, their domain architecture and number of cysteine-rich repeats per protein were defined by jackhmmer in the HMMER 3.1b2 package [74]. The same package was used to automatically assign every individual domain to a specific protein family. Domains were considered conserved when they matched the existing Pfam 32.0 database entries [43, 75]. We hypothesized that taxonomic distribution of a domain might differ from the taxonomic distribution of a protein. Thus, confinement of a conserved domain to a specific taxon was derived from the species distribution information in a Pfam 32.0 database (<http://pfam.xfam.org/>) [43, 75]. A domain was considered typical to prokaryotes or viruses if more than three quarters of carrier species belonged to those groups. The functions of the discovered conserved domains as a possible cell-wall hydrolytic enzyme was inferred based on the information

from Pfam 32.0, InterPro 74.0 (<http://www.ebi.ac.uk/interpro/>) [76] and CAZy 2019/03/20 (<http://www.cazy.org/>) [77] databases, as well as the original literature [64].

Search for homologous sequences

One of the proteins from the fungus *Neocallimastix californiane* (A0A1Y2FMX2) was thoroughly analyzed. Signal peptide position in amino acid sequence was predicted with SignalP5.0 [78]. Intron position in the corresponding genomic sequence was retrieved from the whole genome shotgun sequence of *N. californiane* (GenBank MCOG01000004.1, positions 1,277,192–1,279,411). Positions of the conserved domains were predicted using InterPro 74.0. Finally, amino acid sequences of the individual domains, cut out of the whole protein sequence, were subjected to a remote homology search in the Uniclust30_2018_08 database [79] using HHblits 3.2.0 [80, 81] of the MPI Bioinformatics Toolkit web site (<https://toolkit.tuebingen.mpg.de/>).

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12859-020-03599-y>.

Additional file 1: Supplementary Table T1. Results of three iterations of jackhammer search. Proteins containing cysteine-rich repeats.

Abbreviations

HGT: Horizontal gene transfer; AttP: Bacteriophage recombination site

Acknowledgements

We would like to thank Grants Council of the President of the Russian Federation for the scholarship of the President of Russian Federation for young scientists and PhD students. The authors also appreciate the help received at the Kartesh White Sea Biological Station of the Zoological Institute of the Russian Academy of Sciences. We used the core facilities of the Research Park of St. Petersburg State University: Center for Molecular and Cell Technologies. We would like to express special thanks to Dr. Boto (The National Museum of Natural Sciences, Madrid, Spain) for providing useful comments that improved the manuscript.

About this supplement

This article has been published as part of [BMC Bioinformatics, Volume 21 Supplement 12, 2020: Selected abstracts and papers of Bioinformatics: from Algorithms to Applications 2019 conference. The full contents of the supplement are available at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-21-supplement-12>].

Authors' contributions

MD and SS performed the analysis, OP designed the study. MD drafted the manuscript. Joint efforts went into discussion and editing of the later drafts. The authors read and approved the final manuscript.

Funding

Publication of this supplement was funded by Russian Science Foundation (grant no. 19–74–20102) which provided access to online resources.

Availability of data and materials

All data analyzed during this study are publicly available at GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), UniProtKB (<https://www.uniprot.org/>), Pfam (<https://pfam.xfam.org/>), InterPro (<https://www.ebi.ac.uk/interpro/>) and CAZy (<http://www.cazy.org/>) databases. All data generated during this study are included in this published article [and its supplementary information files].

Ethics approval and consent to participate

'Not applicable'.

Consent for publication

'Not applicable'.

Competing interests

Authors declare no conflict of interests.

Author details

¹Institute of Cytology, St. Petersburg, Russia 194064. ²School of Biomedicine, Far Eastern Federal University, Vladivostok, Russia 690090. ³Department of Cytology and Histology, St. Pb State University, St. Petersburg, Russia 199034.

Received: 8 June 2020 Accepted: 10 June 2020

Published: 24 July 2020

References

- Boto L. Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci.* 2010;277(1683):819–27.
- Wolf YI, Koonin EV. Genome reduction as the dominant mode of evolution. *BioEssays.* 2013;35(9):829–37.
- Domingues S, Nielsen KM. Membrane vesicles and horizontal gene transfer in prokaryotes. *Mob Genet Elem HGT Prokaryotes Microbiota.* 2017;38:16–21.
- Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol.* 2005;3(9):711–21.
- Boto L. Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc R Soc B Biol Sci.* 2014;281(1777):20132450.
- Husnik F, McCutcheon JP. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol.* 2017;16:67.
- Lacroix B, Citovsky V. Transfer of DNA from Bacteria to Eukaryotes. *mBio.* 2016;7(4):e00863–16.
- Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, et al. Designer diatom episomes delivered by bacterial conjugation. *Nat Commun.* 2015;6(1):6925.
- Heinemann JA, Sprague GF. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature.* 1989;340(6230):205–9.
- Waters VL. Conjugation between bacterial and mammalian cells. *Nat Genet.* 2001;29(4):375–6.
- Schönknecht G, Weber APM, Lercher MJ. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *BioEssays.* 2014;36(1):9–20.
- Grau-Bové X, Ruiz-Trillo I, Rodríguez-Pascual F. Origin and evolution of lysyl oxidases. *Sci Rep.* 2015;5:10568.
- Jackson DJ, Macis L, Reitner J, Wörheide G. A horizontal gene transfer supported the evolution of an early metazoan biomineralization strategy. *BMC Evol Biol.* 2011;11(1):238.
- Syvanen M. Evolutionary implications of horizontal gene transfer. *Annu Rev Genet.* 2012;46:341–58.
- Villarreal LP, Witzany G. Viruses are essential agents within the roots and stem of the tree of life. *J Theor Biol.* 2010;262(4):698–710.
- Gilbert C, Chateigner A, Ernenwein L, Barbe V, Bézier A, Herniou EA, et al. Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. *Nat Commun.* 2014;5:3348.
- Filée J, Pouget N, Chandler M. Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses. *BMC Evol Biol.* 2008;8:320.
- Moreira D, Brochier-Armanet C. Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. *BMC Evol Biol.* 2008;8:12.
- Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* 2015;16(1):50.
- Riley DR, Sieber KB, Robinson KM, White JR, Ganesan A, Nourbakhsh S, et al. Bacteria-Human Somatic Cell Lateral Gene Transfer Is Enriched in Cancer Samples. *PLoS Comput Biol.* 2013;9(6):e1003107 Eisen JA, редактор.
- Lehti TA, Pajunen MI, Skog MS, Finne J. Internalization of a polysialic acid-binding Escherichia coli bacteriophage into eukaryotic neuroblastoma cells. *Nat Commun.* 2017;8(1):1–12.
- Redrejo-Rodríguez M, Muñoz-Espín D, Holguera I, Mencía M, Salas M. Functional eukaryotic nuclear localization signals are widespread in terminal proteins of bacteriophages. *Proc Natl Acad Sci.* 2012;109(45):18482.
- Gluck-Thaler E, Slot JC. Dimensions of horizontal gene transfer in eukaryotic microbial pathogens. *PLoS Pathog.* 2015;11(10):e1005156.
- Naranjo-Ortiz MA, Brock M, Brunke S, Hube B, Marcet-Houben M, Gabaldón T. Widespread inter- and intra-domain horizontal gene transfer of d-amino acid metabolism enzymes in eukaryotes. *Front Microbiol.* 2016;7:2001.
- Duarte I, Huynen MA. Contribution of Lateral Gene Transfer to the evolution of the eukaryotic fungus *Piromyces* sp. E2: Massive bacterial transfer of genes involved in carbohydrate metabolism. *bioRxiv.* 2019:514042.
- Nosenko T, Bhattacharya D. Horizontal gene transfer in chromalveolates. *BMC Evol Biol.* 2007;7(1):173.
- Andersson JO. Gene transfer and diversification of microbial eukaryotes. *Annu Rev Microbiol.* 2009;63(1):177–93.
- Andersson JO. Evolution of patchily distributed proteins shared between eukaryotes and prokaryotes: Dictyostelium as a case study. *J Mol Microbiol Biotechnol.* 2011;20(2):83–95.
- Eme L, Gentekaki E, Curtis B, Archibald JM, Roger AJ. Lateral gene transfer in the adaptation of the anaerobic parasite *Blastocystis* to the gut. *Curr Biol.* 2017;27(6):807–20.
- Dunning Hotopp JC. Horizontal gene transfer between bacteria and animals. *Trends Genet TIG.* 2011;27(4):157–63.
- Paganini J, Campan-Fournier A, Rocha MD, Gouret P, Pontarotti P, Wajnborg E, et al. Contribution of Lateral Gene Transfers to the Genome Composition and Parasitic Ability of Root-Knot Nematodes. *PLoS One.* 2012;7(11):e50875.
- Schwarz EM. Evolution: a Parthenogenetic nematode shows how animals become sexless. *Curr Biol.* 2017;27(19):R1064–6.
- Schiffer PH, Danchin EGJ, Burnell AM, Creevey CJ, Wong S, Dix I, et al. Signatures of the Evolution of Parthenogenesis and Cryptobiosis in the Genomes of Panagrolaimid Nematodes. *iScience.* 2019;21:587–602.
- Gladyshev EA, Meselson M, Arkhipova IR. Massive Horizontal Gene Transfer in Bdelloid Rotifers. *Science.* 2008;320(5880):1210.
- Flot J-F, Hespels B, Li X, Noel B, Arkhipova I, Danchin EGJ, et al. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature.* 2013;500:453.
- Nowell RW, Almeida P, Wilson CG, Smith TP, Fontaneto D, Crisp A, et al. Comparative genomics of bdelloid rotifers: insights from desiccating and nondesiccating species. *PLoS Biol.* 2018;16(4):e2004830.
- Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, et al. The dynamic genome of hydra. *Nature.* 2010;464:592.
- Alié A, Hiebert LS, Simion P, Scelzo M, Prünster MM, Lotito S, et al. Convergent Acquisition of Nonembryonic Development in Styelid ascidians. *Mol Biol Evol.* 2018;35(7):1728–43.
- Scelzo M, Alié A, Pagnotta S, Lejeune C, Henry P, Gilletta L, et al. Novel budding mode in *Polyandrocarpa zorritensis*: a model for comparative studies on asexual development and whole body regeneration. *EvoDevo.* 2019;10:7–19.

40. Nakashima K, Yamada L, Satou Y, Azuma J, Satoh N. The evolutionary origin of animal cellulose synthase. *Dev Genes Evol.* 2004;214(2):81–8.
41. Daugavet MA, Shabelnikov S, Shumeev A, Shaposhnikova T, Adonin LS, Podgornaya O. Features of a novel protein, rusticalin, from the ascidian *Styela rustica* reveal ancestral horizontal gene transfer event. *Mob DNA.* 2019;10:4–4.
42. Schuster-Böckler B, Schultz J, Rahmann S. HMM logos for visualization of protein families. *BMC Bioinformatics.* 2004;5:7.
43. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* 2019;47(Database issue):D427–32.
44. Metcalf JA, Funkhouser-Jones LJ, Briley K, Reysenbach A-L, Bordenstein SR. Antibacterial gene transfer across the tree of life. *eLife.* 2014;3:e04266.
45. Moran Y, Fredman D, Szczesny P, Grynberg M, Technau U. Recurrent horizontal transfer of bacterial toxin genes to eukaryotes. *Mol Biol Evol.* 2012;29(9):2223–30.
46. Klasson L, Kambris Z, Cook PE, Walker T, Sinkins SP. Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. *BMC Genomics.* 2009;10:33.
47. White SH, Wimley WC, Selsted ME. Structure, function, and membrane integration of defensins. *Curr Opin Struct Biol.* 1995;5(4):521–7.
48. Ding J, Chou Y-Y, Chang TL. Defensins in viral infections. *J Innate Immun.* 2009;1(5):413–20.
49. Wilson SS, Wiens ME, Smith JG. Antiviral Mechanisms of Human Defensins. *J Mol Biol.* 2013;425(24). <https://doi.org/10.1016/j.jmb.2013.09.038>.
50. Silva J, Loreto E, Clark J. Factors that affect horizontal transfer of transposable elements. *Curr Issues Mol Biol.* 2004;6:57–71.
51. Lespinet O, Wolf YI, Koonin E, Aravind L. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* 2002;12:1048–59.
52. Schilbert HM, Pellegrinelli V, Rodríguez-Cuenca S, Vidal-Puig A, Pucker B. Harnessing natural diversity to identify key residues in Prolidase. *bioRxiv.* 2018:423475.
53. Ochoa A, Storey JD, Llinás M, Singh M. Beyond the E-value: stratified statistics for protein domain prediction. *PLoS Comput Biol.* 2015;11(11):e1004509.
54. Wang G, Liu Z, Lin R, Li E, Mao Z, Ling J, et al. Biosynthesis of antibiotic Leucinostatins in bio-control fungus *Purpureocillium lilacinum* and their inhibition on *Phytophthora* revealed by genome mining. *PLoS Pathog.* 2016;12(7):e1005685.
55. Marcet-Houben M, Gabaldón T. Acquisition of prokaryotic genes by fungal genomes. *Trends Genet TIG.* 2009;26:5–8.
56. Hu X, Xiao G, Zheng P, Shang Y, Su Y, Zhang X, et al. Trajectory and genomic determinants of fungal-pathogen speciation and host adaptation. *Proc Natl Acad Sci U S A.* 2014;111(47):16796–801.
57. Ren Q, Wang C, Jin M, Lan J, Ye T, Hui K, et al. Co-option of bacteriophage lysozyme genes by bivalve genomes. *Open Biol.* 2017;7(1):160285.
58. Chan CX, Darling AE, Beiko RG, Ragan MA. Are protein domains modules of lateral genetic transfer? *PLoS One.* 2009;4(2):e4524.
59. Ragan MA, Beiko RG. Lateral genetic transfer: open issues. *Philos Trans R Soc Lond Ser B Biol Sci.* 2009;364(1527):2241–51.
60. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol.* 2018;36:996.
61. Bochtler M, Odintsov SG, Marcyjaniak M, Sabala I. Similar active sites in lysostaphins and D-Ala-D-Ala metallopeptidases. *Protein Sci.* 2004;13(4):854–61.
62. Vollmer W, Joris B, Charlier P, Foster S. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol Rev.* 2008;32(2):259–86.
63. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* 2008;37(suppl_1):D233–8.
64. Vermassen A, Leroy S, Talon R, Provot C, Popowska M, Desvaux M. Cell Wall hydrolases in Bacteria: insight on the diversity of Cell Wall Amidases, Glycosidases and Peptidases Toward Peptidoglycan. *Front Microbiol.* 2019;10:331.
65. Degnan S. Think laterally: horizontal gene transfer from symbiotic microbes may extend the phenotype of marine sessile hosts. *Front Microbiol.* 2014;5:638.
66. Hirt RP, Alsmark C, Embley TM. Lateral gene transfers and the origins of the eukaryote proteome: a view from microbial parasites. *Curr Opin Microbiol.* 2015;23:155–62.
67. Fernandes S, São-José C. Enzymes and mechanisms employed by tailed bacteriophages to breach the bacterial cell barriers. *Viruses.* 2018;10(8):396.
68. Jollès P. *Lysozymes: Model Enzymes in Biochemistry and Biology.* NY: Springer Verlag; 1996. p. 478.
69. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. *J Virol.* 2010;84(19):9733.
70. Domingo-Calap P, Schubert B, Joly M, Solis M, Untrau M, Carapito R, et al. An unusually high substitution rate in transplant-associated BK polyomavirus in vivo is further concentrated in HLA-C-bound viral peptides. *PLoS Pathog.* 2018;14(10):e1007368.
71. Koonin E, Galperin MY. *Sequence — evolution — function: computational approaches in comparative genomics.* Springer US; 2003.
72. George RA, Heringa J. The REPRO server: finding protein internal sequence repeats through the web. *Trends Biochem Sci.* 2000;25(10):515–7.
73. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
74. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 update. *Nucleic Acids Res.* 2015;43(W1):W30–8.
75. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016;44(D1):D279–85.
76. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* 2017;45(D1):D190–9.
77. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014;42(Database issue):D490–5.

78. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019;37(4):420–3.
79. Mirdita M, von den Driesch L, Galiez C, Martin MJ, Söding J, Steinegger M. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* 2016;45(D1):D170–6.
80. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods.* 2011;9:173.
81. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, et al. A completely Reimplemented MPI bioinformatics toolkit with a new HHpred server at its Core. *Comput Resour Mol Biol.* 2018;430(15):2237–43.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

