

RESEARCH ARTICLE

Open Access

A manual collection of *Syt*, *Esy*, *Rph3a*, *Rph3al*, *Doc2*, and *Dblc2* genes from 46 metazoan genomes - an open access resource for neuroscience and evolutionary biology

Molly Craxton*

Abstract

Background: Synaptotagmin proteins were first identified in nervous tissue, residing in synaptic vesicles. Synaptotagmins were subsequently found to form a large family, some members of which play important roles in calcium triggered exocytic events. These members have been investigated intensively, but other family members are not well understood, making it difficult to grasp the meaning of family membership in functional terms. Further difficulty arises as families are defined quite legitimately in different ways: by common descent or by common possession of distinguishing features. One definition does not necessarily imply the other. The evolutionary range of genome sequences now available, can shed more light on synaptotagmin gene phylogeny and clarify family relationships. The aim of compiling this open access collection of synaptotagmin and synaptotagmin-like sequences, is that its use may lead to greater understanding of the biological function of these proteins in an evolutionary context.

Results: 46 metazoan genomes were examined and their complement of *Syt*, *Esy*, *Rph3a*, *Rph3al*, *Doc2* and *Dblc2* genes identified. All of the sequences were compared, named, then examined in detail. *Esy* genes were formerly named *Fam62*. The species in this collection are *Trichoplax*, *Nematostella*, *Capitella*, *Helobdella*, *Lottia*, *Ciona*, *Strongylocentrotus*, *Branchiostoma*, *Ixodes*, *Daphnia*, *Acyrtosiphon*, *Tribolium*, *Nasonia*, *Apis*, *Anopheles*, *Drosophila*, *Caenorhabditis*, *Takifugu*, *Tetraodon*, *Gasterosteus*, *Oryzias*, *Danio*, *Xenopus*, *Anolis*, *Gallus*, *Taeniopygia*, *Ornithorhynchus*, *Monodelphis*, *Mus* and *Homo*. All of the data described in this paper is available as additional files.

Conclusions: Only a subset of synaptotagmin proteins appear able to function as calcium triggers. *Syt1*, *Syt7* and *Syt9* are ancient conserved synaptotagmins of this type. Some animals carry extensive repertoires of synaptotagmin genes. Other animals of no less complexity, carry only a small repertoire. Current understanding does not explain why this is so. The biological roles of many synaptotagmins remain to be understood. This collection of genes offers prospects for fruitful speculation about the functional roles of the synaptotagmin repertoires of different animals and includes a great range of biological complexity. With reference to this gene collection, functional relationships among *Syt*, *Esy*, *Rph3a*, *Rph3al*, *Doc2* and *Dblc2* genes, which encode similar proteins, can better be assessed in future.

Background

Synaptotagmin (*Syt*) proteins participate in regulated membrane fusion events in multicellular organisms. *Syt* research dates back to 1981, when the first *Syt* was identified as an integral protein of synaptic vesicles [1]. Since then, much effort has gone into discovering the

function and the detailed mechanism of action of this protein, *Syt1*. Numerous experimental approaches have shown that *Syt1* is crucially involved in fast neurotransmitter release at synapses [reviewed in [2-6]]. *Syt1* binds calcium, serving as the calcium sensor which triggers synaptic vesicle exocytosis. *Syt1* also binds to the neuronal SNARE proteins which are required for membrane fusion, as well as to membranes directly. The primary structure of *Syt1* [7] revealed three important features: an N-terminal transmembrane (TM) domain, which

* Correspondence: mollycraxton@btinternet.com
Medical Research Council Laboratory of Molecular Biology, Hills Road,
Cambridge, CB2 0QH, UK

serves to anchor the protein in the synaptic vesicle, plus two, tandem, C-terminal, cytoplasmic C2 domains (C2A and C2B) which specify the calcium, SNARE and membrane binding properties. Crystallographic studies of some Syt C2 domains [8,9] have revealed the exact nature of their calcium binding abilities. Structural bioinformatics [10-12] has shown that C2 domains are very common in the eukaryotic protein repertoire, but not all act to bind calcium. C2 domains can occur singly or as multiple copies in a given protein [13]. The particular domain organisation of Syt1 is important for its role in synaptic vesicle exocytosis. The TM domain tethers the protein to the membrane. The C2A and C2B domains, endowed with individual capacities to bind calcium, SNARE proteins, other accessory proteins and phospholipids, act both independently and synergistically, at different stages during the life cycle of the synaptic vesicle, to promote or inhibit fusion [14-16].

After the genetic code for Syt1 was discovered [7] genes similar to *Syt1* were sought and found. Biochemical methods which were employed initially, led to an increase in the size of the rodent *Syt* family, from one member to thirteen members and beyond. *Syt* relatives were also identified and studied in *D. melanogaster* and *C. elegans*. All of these homologous *Syt* genes were found to encode proteins with a common domain architecture: an N-terminal TM domain connected by a variable length, poorly conserved linker sequence, to well conserved, tandem, C-terminal C2A and C2B domains and this domain architecture was taken to define the family. Functional studies with the proteins *in vivo* and *in vitro*, indicate that some respond to calcium but others do not. Whether or not they respond to calcium, Syt proteins are able to regulate membrane fusion due to common properties of their C2 domains [17-21]. Studies of the anatomical expression patterns of *Syt* genes show that *Syt1* is abundantly expressed in nervous tissue [1,22-26]. In *D. melanogaster*, *Syt1* expression is neuron specific but other *Syt* genes are expressed elsewhere [27]. In rodents, most *Syt* genes are expressed in the brain [28].

With the publication of whole genome sequences from multicellular organisms, it became possible to identify and compare complete genome complements of *Syt* genes using computational methods [29,30]. The phylogeny of these genes could be examined by comparing sequences from different organisms. Seventeen *Syt* homologues were found in each of the complete *H. sapiens* and *M. musculus* genome sequences. These were named *SYT1* (*Syt1*) to *SYT17* (*Syt17*) according to the nomenclature conventions of the HUGO Gene Nomenclature Committee [31] (and the Mouse Genome Informatics Database [32]). Some of these genes however, encode Syt proteins which lack the requisite

domain architecture for inclusion in a Syt family defined by domain structure. The complete genome sequences of *C. elegans*, *D. melanogaster* and *A. thaliana* were analysed, together with draft genome sequences from other eukaryotes, to assess their Syt coding potential [30]. Proteins which share the stereotypical Syt domain architecture, are encoded by distinct (not phylogenetically homologous) gene families in plants and in animals [33] further confounding the notion of a homologous Syt family defined in terms of domain architecture. Functional similarity between plant and animal Syt proteins, at the level of calcium and membrane binding, has been demonstrated [34,35] in keeping with the notion that protein form specifies function. Proteins have been classified as Syt family members by the presence or absence of suitable characteristics in terms of protein form and function [eg. [6,36,37]] without studious regard to gene ancestry. Legitimately but confusingly, family membership can be defined in different ways: in terms of descent from a common ancestor (homology) or in terms of the common possession of distinguishing features. Since it is not straightforward to demonstrate that homology is the cause of the similarity between biological entities such as proteins, it is often just assumed, leading to the confusion of two different notions of a protein family. Please see [38,39] for explanations of the terminology of homology. Examination of the genomic specification of eukaryotic proteins can help to reveal their phylogenetic relationships because the evolutionary conservation of gene structure provides additional information on which to base an inference of homology. Families may thereby be resolved as those related by homology and those otherwise related. Ambiguity still remains however, because depending on the exact biological context, members of a family, however defined, can sometimes be considered as functionally interchangeable and thus worthy of a common functional identity. Thus, genes and proteins can acquire multiple identities depending on the perspective from which they are considered. Universal agreement about nomenclature for genes and proteins which is suitably clear and meaningful is a challenge for the future, but norms for gene nomenclature according to phylogenetic relationships are established [31,32].

It may be the case (as is often assumed for an orthologous family group) that members of a gene family related by heredity, share a common functional role [38-41]. However, due to the dynamics of eukaryotic genome evolution [41-45] the functional attributes of gene relatives can diverge. It is also possible, that separate start points, through advantageous gain of function, can lead through selection, to convergence upon a common form and function. While it is practicable to assign family membership to genes or proteins based on their

apparent hereditary relationships, the functional implications of family membership cannot simply be deduced. More than a decade of research effort has been aimed at understanding how Syt1 works. Considerable effort has also been directed at understanding the functions of other Syt proteins. Current understanding however, is insufficient to predict the functions of a whole genome complement of Syt proteins, or indeed to predict what a genome complement of Syt proteins might consist of.

To help improve this lack of understanding, I have put together an open access resource intended to serve future research aimed at understanding the biological meaning of the hereditary and functional relationships among some of the metazoan Syt-like proteins. I have collected and compiled manually, information about homologous *Syt*, *Esy*, *Rph3a*, *Rph3al*, *Doc2* and *Dblc2* genes from 46 metazoan genome sequences spanning a wide evolutionary range. I have not attempted to collect all genes capable of encoding proteins with similarity on any level to Syt proteins. Such a collection would ultimately include perhaps, all eukaryotic C2 domain proteins. My collection is restricted to homologous *Syt* genes plus a small group of genes already identified as similar but not homologous to *Syt* genes [33]. The collection does include a new group of homologous genes (*Dblc2*) which are present in the genomes of marine invertebrates and encode proteins similar to Syt proteins. The information in this collection may interest investigators in the field of evolutionary biology, with examples of dynamic genome evolution including whole gene duplication, partial duplication, gene fission, acquisition of novel coding exons, gene inactivation, extreme sequence divergence, intron mobilization and a variety of routes to the expression of altered gene products. For experimentalists who want to understand how Syt proteins act to effect the biology of animal nervous systems, the collection provides a utility with great evolutionary depth. In order to make progress in neuroscience, appropriate model organisms must be chosen. The choice of model organism is fundamental to the type of science which can be pursued. The range of organisms and genes in this collection, offer novel possibilities for future descriptive, comparative and hypothesis driven research.

In order to make the large amount of information described by this paper available for public scrutiny, it is available as 50 additional files which can be accessed online.

Results and Discussion

Origin of this gene collection

The origin of this manual gene curation project lay in an attempt to annotate some of the plant *NTMC2* genes I had identified previously [33]. I chose the unique

NTMC2 nomenclature to emphasise the phylogenetic relationships among these plant genes and to distinguish them from animal *Syt* genes. Plant *NTMC2T1*, *NTMC2T2* and *NTMC2T3* genes resemble animal *Syt* genes in having the same domain architecture. That this shared domain architecture implies functions in common, has now been demonstrated experimentally [34,35]. The *A. thaliana* *NTMC2T1.1* protein (also known as SYT1 [34,35,37]) possesses calcium and membrane binding activities which allow it to function in a plasma membrane repair pathway induced by stress. While using the annotation facilities at DOE-JGI [46] I realised that the evolutionary range of genome sequences available for inspection at DOE-JGI, could be used to search for earlier origins of the embryophyte *NTMC2* genes and metazoan *Syt* genes. The genome sequence of *M. brevicollis* [47] has demonstrated that, as theorised [48,49] the evolutionary transition from unicellularity to multicellularity involved abundant domain shuffling in proteins. *M. brevicollis*, a unicellular protist, closely related but basal to metazoans, does not possess *Syt* genes. It does have a gene with a degree of gene structure similarity, domain architecture similarity and amino acid sequence similarity to plant *NTMC2* and animal *Esy* proteins XM_001748216. Proteins with similarity to *NTMC2* and *Esy* proteins (but without similarity at the gene structure level) are present in other unicellular eukaryotes [33]. It seems reasonable therefore, to propose that while *NTMC2* and *Esy* genes could ultimately be homologous, having attained their current forms through descent from a unicellular ancestor, *Syt* genes, distinguished by their conserved gene structure, more likely represent a metazoan novelty.

On finding uniquely metazoan *Syt* homologues, with their characteristic gene structures, in the genomes of *T. adhaerens* and *N. vectensis* [50,51] I decided to compile afresh, the *Syt* repertoire of the currently available genomes of evolutionary significance. In addition to *Syt* genes, I set out to gather *Esy*, *Rph3a*, *Rph3al* and *Doc2* homologues. These gene families, while clearly distinct from the *Syt* family at the gene structure level, encode proteins with high level sequence similarity to the tandem C2 domain region of *Syt* (*Rph3a* and *Doc2*) or with similarity, in common with *NTMC2* proteins, at a more crude domain architecture level (*Esy*). Because of these similarities in form, there will be common functions. A collection like this can be used to further define these functions, by revealing the patterns of sequence conservation (indicating selection and functional importance) and sequence divergence, apparent in: 1. the individual domains of an orthologous group of proteins, 2. the equivalent domain in different groups, not necessarily homologous. Such knowledge can provide clues about how these proteins function as molecular machines in

cells and about how these functions have developed over evolutionary time. I used the annotation facilities at DOE-JGI to create and annotate gene models where possible. Full details of all genes are listed in additional files 1, 2, 3 and 4. Figures 1, 2 and 3 summarise the genes in additional files 1, 2 and 3 respectively.

Gene identification

I searched for homologues among the metazoan genome sequences at DOE-JGI [46] NCBI [52] and UCSC [53]. I employed a variety of methods, first to identify, then to piece together, gene sequences for homologues of *Syt*, *Esy*, *Rph3a*, *Rph3al* and *Doc2*. These are generally large, multi-exon genes which encompass regions of high (the C2 domain region of Syt proteins for example) and low (the N-terminal portion of Syt proteins for example) levels of sequence conservation. As a first step, I used blatp [54] or tblastn [55] with amino acid sequence probes representing already known homologues, to identify genomic loci with potential to encode a similar protein. I then examined a 3-frame translation [56] of the genomic region to look at the possible exon-intron structure. Intron positions and phases are generally well conserved among these genes and serve as a useful diagnostic tool. In cases where the draft genome sequence was highly fragmented, homologues were pieced together exon by exon. Where sequence conservation between amino acid probe and novel genomic locus was high, it was straightforward to identify exons and introns. Where sequence conservation was low, the conservation of intron position and phase could serve as a guide to judge possible exon boundaries. In all cases, when transcript sequences from the locus under examination, or from a homologous locus in a closely related

organism, were available, these were used to help identify exons. Occasionally, trace archive reads were sought to confirm exon sequences, where there was a gap or where there appeared to be frameshift errors in a section of draft sequence for example. Details of such instances are noted in additional files 1, 2 and 3.

Where transcript sequences indicated the expression of alternative variants, I collected the variants. In the case of *M. musculus* and *H. sapiens*, where there is abundant transcript variation, particularly at certain genomic loci, I collected only those variants indicated by two or more transcript sequences. Where evidence of alternative splicing exists but is not represented by full length transcripts (the majority of cases) the resultant gene predictions require validation, as do all gene predictions for which no transcript evidence exists. For genomic loci with conserved variant exons, which could be employed to extend the length of the encoded protein but where transcript evidence was lacking, the gene prediction includes all of the conserved exons. Transcript evidence for the gene predictions is listed in additional files 1, 2, 3 and 4.

Once I had gathered a set of crude gene predictions, I compared the sequences in order to refine the predictions. Exon-intron junctions, gaps, novel regions of sequence conservation useful as probes to fill gaps, were all carefully examined using multiple alignments [57]. In this way, it was possible to extend and improve the gene predictions significantly. Many of the gene predictions in this collection are still incomplete, at the N-terminal region of Syt proteins in particular. When there was no transcript coverage and no detectable sequence conservation or very short coding exons, it was not possible to

Marine Invertebrates									
<i>Trichoplax adhaerens</i>	<i>Nematostella vectensis</i>	<i>Capitella sp. 1</i>	<i>Helobdella robusta</i>	<i>Lottia gigantea</i>	<i>Ciona savignyi</i>	<i>Ciona intestinalis</i>	<i>Stongylocentrotus purpuratus</i>	<i>Branchiostoma floridae</i>	
http://www.jgi.doe.gov	http://www.broad.mit.edu	http://www.jgi.doe.gov	http://www.hgsc.bcm.tmc.edu	http://www.jgi.doe.gov					
Gene Name	Gene Name	Gene Name	Gene Name	Gene Name					
Sy1	Sy1	Sy1	Sy1a	Sy1	Sy1	Sy1	Sy1	Sy1a	Sy1a
Sy7	Sy7	Sy4	Sy1b - sequence starts within C2AB region	Sy4	Sy7	Sy7	Sy4	Sy7	Sy1b
Sy21	Sy21	Sy7	Sy1c	Sy7	Sy15	Sy15	Sy7	Sy7	Sy4
Sy22	Sy24	Sy9	Sy1d	Sy9	Sy16	Sy16	Sy9a	Sy7	Sy7
Sy23	Sy25	Sy12	Sy1e	Sy12	Sy16	Sy16	Sy9b	Sy9	Sy9
Dblc2	Sy15	Sy15	Sy7a	Sy15a	Sy16	Sy16	Sy12	Sy12	Sy115a
Esy2	Sy27	Sy16	Sy7b	Sy15b	Dblc2	Dblc2	Sy15	Sy115b	Sy115b
Rph	Sy28	Sy17	Sy7c	Sy16	Esy2	Esy2	Sy16	Sy117	Sy117
	Sy29	Sy44	Sy7d	Sy17	Rph	Rph	Sy17	Sy118	Sy118
	Sy30	Sy1alpha	Sy7e	Sy18			Sy18	Sy118	Sy118
	Sy31	Dblc2	Sy7f	Sy21			Sy21	Sy48	Sy152
	Sy32	Esy2	Sy7f	Sy47			Sy48	Sy49	Sy153
	Sy33	Rph	Sy115a	Sy1alpha			Sy49	Sy50	Sy154
	Sy34		Sy115b	Dblc2			Sy50	Sy51	Sy1alpha
	Sy35		Sy116	Esy2			Sy51	Dblc2	Esy2a
	Sy36		Sy45	Rph			Sy1alpha	Esy2	Esy2b
	Sy37		Sy46				Dblc2	Rph - gap in genome sequence	Rph
	Sy38		Esy2a				Esy2		
	Sy39		Esy2b				Rph		
	Sy40		Rph						
	Sy41								
	Sy42								
	Sy43								
	Dblc2								
	Esy2a								
	Esy2b								
	Esy2c								
	Rph								

Figure 1 Summary of the genes collected from marine invertebrate genomes. The website of the organisation which sequenced the genome is listed below the organism name. Underneath the Gene Name heading, gene symbols are listed. Red symbols indicate sequences containing all ten acidic amino acid positions required for function as a calcium trigger for exocytosis.

Ecdysozoa																							
<p><i>Ixodes scapularis</i> http://www.jcvi.org http://www.broad.mit.edu</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt9 Syt13 Syt15a Syt15b Syt16 Syt55 Esy2 Rph - gap in genome sequence Rph1</p>	<p><i>Daphnia pulex</i> http://www.jgi.doe.gov</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt13 Syt15 Syt16 Syt17 Syt56 Syt57 Esy2 Rph1</p>	<p><i>Acyrtosiphon pisum</i> http://www.hgsc.bcm.tmc.edu</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Esy2</p>	<p><i>Tribolium castaneum</i> http://www.hgsc.bcm.tmc.edu</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt9 Syt12 Syt13 Syt15 Syt16 Esy2a Esy2b Esy2c Rph</p>	<p><i>Nasonia vitripennis</i> http://www.hgsc.bcm.tmc.edu</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt9 Syt12 Syt13 Syt16 Syt20 Sytalpa Esy2</p>	<p><i>Apis mellifera</i> http://www.hgsc.bcm.tmc.edu</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt9 Syt12 Syt13 Syt16 Syt20 Esy2 Rph</p>	<p><i>Anopheles gambiae</i> http://www.jcvi.org http://genome.wustl.edu</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt13 Syt16 Sytalpa Esy2</p>	<p><i>Drosophila melanogaster</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila simulans</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila sechellia</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila erecta</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila yakuba</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila ananassae</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila pseudoobscura</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila persimilis</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila willistoni</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila virilis</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila mojavensis</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Drosophila grimshawi</i> http://flybase.org</p> <p>Gene Name</p> <p>Syt1 Syt4 Syt7 Syt12 Syt13 (Sytbeta) Syt16 (Syt14) Sytalpa Esy2 (G6643) Rph</p>	<p><i>Caenorhabditis elegans</i> http://genome.wustl.edu http://www.wormbase.org</p> <p>Gene Name</p> <p>snt-1 snt-4 snt-6 snt-2 snt-3 snt-5 snt-7 esyf-2 rbf-1</p>	<p><i>Caenorhabditis brenneri</i> http://genome.wustl.edu</p> <p>Gene Name</p> <p>snt-1 snt-4 snt-6 snt-2 snt-3 snt-5 snt-7 esyf-2 rbf-1</p>	<p><i>Caenorhabditis briggsae</i> http://genome.wustl.edu</p> <p>Gene Name</p> <p>snt-1 snt-4 snt-6 snt-2 snt-3 snt-5 snt-7 esyf-2 rbf-1</p>	<p><i>Caenorhabditis remanei</i> http://genome.wustl.edu</p> <p>Gene Name</p> <p>snt-1 snt-4 snt-6 snt-2 snt-3 snt-5 snt-7 esyf-2 rbf-1</p>	<p><i>Caenorhabditis japonica</i> http://genome.wustl.edu</p> <p>Gene Name</p> <p>snt-1 snt-4 snt-6 snt-2 snt-3 snt-5 snt-7 esyf-2 rbf-1</p>

Figure 2 Summary of the genes collected from ecdysozoan genomes. The websites of the organisations which sequenced the genome or which provide access to multiple genomes within a single genus, are listed below the organism name. Underneath the Gene Name heading, gene symbols are listed. Red symbols indicate sequences containing all ten acidic amino acid positions required for function as a calcium trigger for exocytosis. Websites for the relevant nomenclature authorities are listed alongside the Gene Name heading. Gene symbols within brackets are currently officially approved, but in conflict with the nomenclature proposed here.

make a prediction. All incomplete predictions are noted in additional files 1, 2, 3 and 4.

Sequence comparison: global comparison and nomenclature

Having collected 711 genes from 46 genomes, I used clustalw2 at EBI [58,59] to compare subsets of amino acid sequences, in order to examine their relationships of similarity overall. I compared sets which excluded expressed variants but which included one sequence per gene, whether complete or not. Comparisons of invertebrate sequences (additional file 5) vertebrate sequences (additional file 6) invertebrate (one representative per genus) plus mammalian sequences (additional file 7) and invertebrate (one representative per genus) plus a subset of vertebrate sequences (additional file 8) were used to examine relationships and assign a suitable name to each gene. I assigned a common name to those genes which clustered together in conserved groups. When a conserved group included a homologue from *M. musculus/H. sapiens*, the assigned name was chosen to be consistent with that in *M. musculus/H. sapiens*. For those *Syt* genes which did not consistently cluster together and are not conserved among different organisms, I assigned them numbers upward from the last conserved group. This process led to the identification of 22 conserved groups of *Syt* genes plus a further 35 unique *Syt* genes named with the *Syt* stem symbol plus

a number identifier, in order to provide a recognisable name (albeit provisional) for each. This naming system allows further sampling of the tree of life by gene sequencing to reveal: 1. more unique *Syt* genes, from *Syt58* upwards, all members of a recognisably homologous family and 2. previously unrecognisable patterns of conservation among *Syt* genes, from *Syt22* upwards. I contacted the relevant organism nomenclature authorities [31,60-66] to discuss this nomenclature. This wide consultation led to agreement to change *FAM62*, a temporary HGNC symbol for this poorly characterised gene family, to *Esyf* [67]. Where current approved nomenclature is discordant with that proposed here, the approved nomenclature is indicated within brackets, in additional files 1, 2, 3 and figures 1, 2 and 3. There is inconsistency in the application of my naming rationale in *Drosophila* and *Caenorhabditis*, where it was more practical to stick with the approved nomenclature than propose new names. As yet, and in comparison to vertebrates, invertebrates are poorly represented by organised gene nomenclature authorities. My proposed nomenclature for a large number of genes in this collection, therefore, awaits sanction. Note that due to differing nomenclature conventions, the gene symbols in different organisms, range from having no capital letters to all capital letters. The generic form, which I have used to label the genes from organisms outside the purview of nomenclature

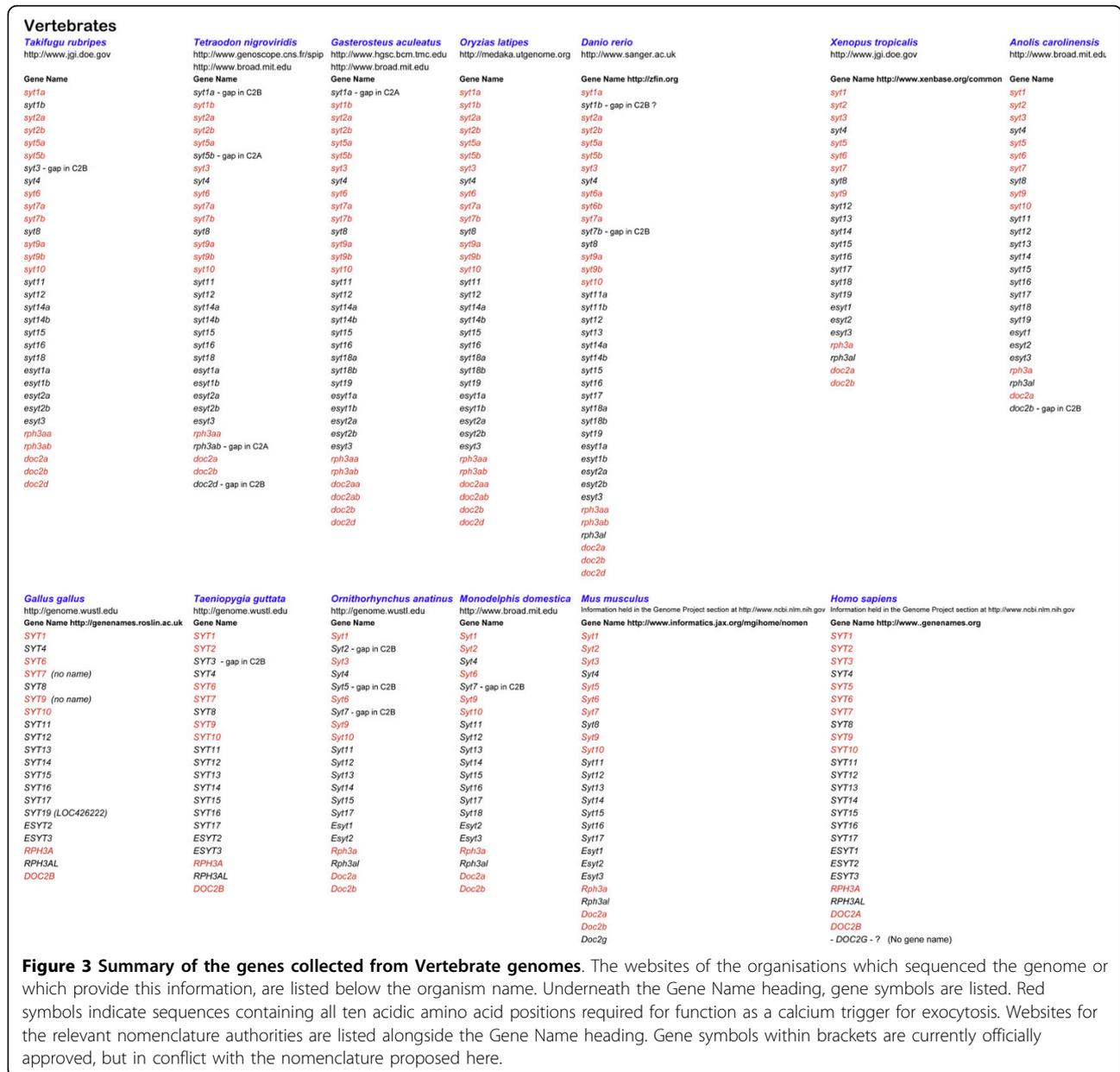


Figure 3 Summary of the genes collected from Vertebrate genomes. The websites of the organisations which sequenced the genome or which provide this information, are listed below the organism name. Underneath the Gene Name heading, gene symbols are listed. Red symbols indicate sequences containing all ten acidic amino acid positions required for function as a calcium trigger for exocytosis. Websites for the relevant nomenclature authorities are listed alongside the Gene Name heading. Gene symbols within brackets are currently officially approved, but in conflict with the nomenclature proposed here.

authorities, is first letter in capitals. This designation will be used here, to describe the gene in general, without reference to a specific organism.

Deciding whether relationships of orthology or some other kind exist among a group of similar genes, can be complicated. The processes of duplication, deletion, and rearrangement of genes, plus the action of selection or absence of selection on gene sequences, can obscure relationships. Whether absence from a genome assembly reflects true absence, or incompleteness of the assembly, presents a further problem of some seriousness [68]. In *M. musculus* and *H. sapiens* there are several groups of duplicated *Syt* genes: group 1 [*Syt1*, *Syt2*, *Syt5*, *Syt8*]

group 2 [*Syt4*, *Syt11*] group 3 [*Syt3*, *Syt6*, *Syt9*, *Syt10*] and group 4 [*Syt14*, *Syt16*]. Each group contains genes which share a distinct gene structure. Additional files 5, 6, 7, 8 and figure 4, indicate that the parental gene of group 1 is *Syt1*. The parental gene of group 2 is *Syt4*, that of group 3 is *Syt9* and that of group 4 is *Syt16*. In each of these cases, the

trying to determine the phylogenetic relationships between the *Syt* genes of mammals and those of other animals because it is the rodent *Syt* proteins which are best characterised functionally. However, because mammalian organisms and their brains in particular, represent biological systems of such complexity, modelling these systems by using simpler systems is essential. The relationships and nomenclature proposed here, are in good general agreement with the *Syt* data at Treefam [69,70] which does not include many of the invertebrate *Syt* genes in this collection.

This gene collection includes several previously unrecognised *Syt* groups which are absent from *M. musculus*/*H. sapiens*. Members of the *Syt18* group appear in vertebrates and invertebrates, but the *Syt19* group has yet to be found beyond vertebrates. Neither of these groups are well conserved, both genes having diverged significantly over time. The *Syt18* group ranges from members encoding: tandem C2 domains (*S. purpuratus*, *B. floridae*, *A. carolinensis*) a single C2B domain (*L. gigantea*, *X. tropicalis*, the fish genomes) or no C2 domains at all (the *Syt18b* duplicates in the fish genomes). *Syt20* appears in only two hexapod species in this collection. The other *Syt* groups absent in vertebrates, are *Sytalpa* and *Syt21*. Of these, the *Sytalpa* group has a higher level of sequence conservation. The *Sytalpa* designation applies to all members of the conserved group which contains the *D. melanogaster Sytalpa* gene (see additional file 5). There are many invertebrate *Syt* genes, related to some degree, at the gene structure level, which are not conserved among different phyla (*Syt22*

to *Syt57* in this collection). *N. vectensis*, for example, has a very large number of unique and divergent *Syt* genes (*Syt24* to *Syt43*). Most of these encode stereotypical *Syt* proteins, but some have only one C2 domain and some lack a TM domain (as noted in additional file 1). The *Caenorhabditis Syt* repertoire, in common with many others in this collection, demonstrates that variant *Syt* proteins which lack the TM domain, are usually included. They are encoded either by separate genes or by alternative splicing. *Caenorhabditis snt-3* is very similar to *snt-1* at the amino acid sequence level, but does not encode a TM domain. In the *Ciona Syt* repertoire, it is *Syt16* which produces a variant lacking the TM domain, but by alternative splicing. In mammals, *Syt17* does not encode a TM domain, but a number of other mammalian *Syt* genes which do encode a TM domain, express splice variants which lack the TM domain (*Syt5*, *Syt6*, *Syt8*, *Syt9*, *Syt10*, *Syt12*, *Syt13*, *Syt15*, *Syt16*). Transcript variants are listed in additional files 1, 2, and 3.

Sequence comparison: gene duplications

The duplicate genes in the fish genomes presented a nomenclature challenge and I am very grateful for help from the Zfin nomenclature coordinator [62,63] in choosing the most appropriate names for these genes based on gene structure similarity and synteny. A number of genes present as a single copy in mammals (*Syt7*, *Syt9*, *Syt14*, *Esy2*, *Rph3a*) are present as duplicates in the fish genomes. In addition, *Syt6* and *Syt11* are duplicated in *D. rerio* (figure 3). It is notable that the single *Rph3a* of tetrapods appears to have undergone duplication and fission, giving rise to *Rph3al* and *Doc2b* genes

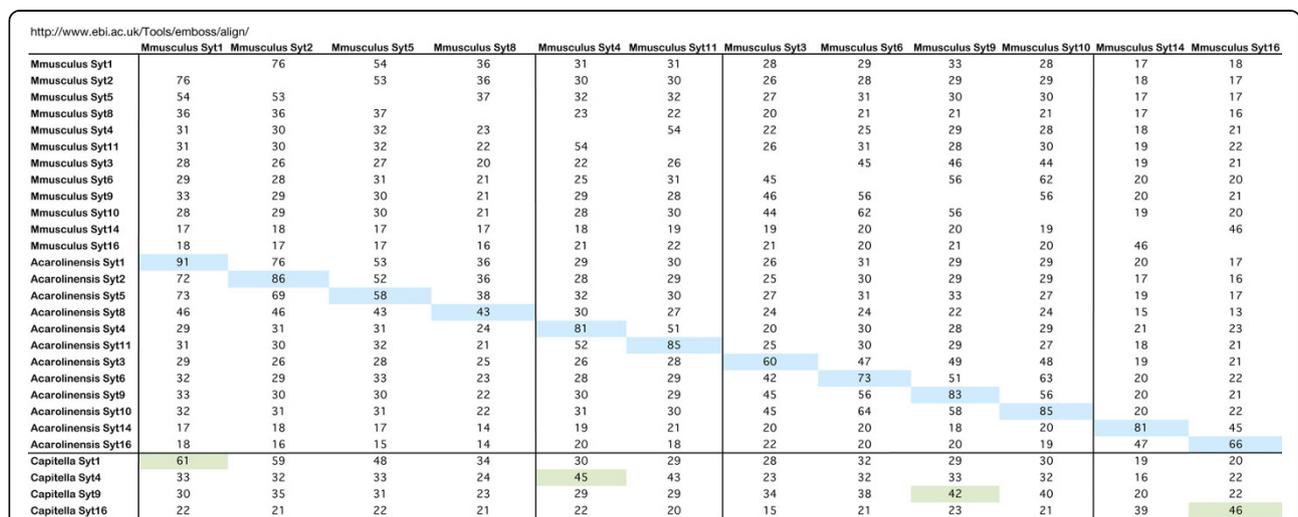


Figure 4 Syt orthologues and paralogues in *M. musculus*. Percent identity scores produced by the align facility at EBI, of pairwise comparisons of full length protein sequences, are listed. Top scores from mouse versus lizard comparisons are highlighted in blue, indicating an orthologous relationship between the mouse gene and the evolutionarily more ancient lizard gene. Top scores from comparisons between mouse and the much more evolutionarily ancient polychaete worm, *Capitella*, are highlighted in green, indicating that of these mouse genes, *Syt1*, *Syt4*, *Syt9* and *Syt16* are orthologous to genes in *Capitella*.

[as noted in 33]. *Rph3al* and *Doc2b* are linked as a tandem pair in tetrapod genomes (noted in additional file 3). Invertebrates have a single, equivalent *Rph*, but no *Doc2* genes. It is possible that *D. pulex Rphl* represents an N-terminal *Rph* fission product. In the fish genomes, there are duplicate *Rph3a* genes, but only in the lineage leading to *D. rerio*, has a supposed *Rph3a* duplication led to the retention of linked *rph3al* and *doc2b* fission products (noted in additional file 3). In the other fish genomes, *doc2b* genes are present, along with the paralogous *doc2a* and *doc2d* genes, as well as duplicate *rph3aa* and *rph3ab* genes, so *Doc2b* genes are not always linked to a *Rph3al* fission product. Apart from the gene duplications already noted, other paralogous gene expansions are evident in this collection (figures 1, 2 and 3). In the case of *Esy1* gene duplications, retro-position of processed transcripts has occasionally led to gene duplication (*N. vectensis Esy1a* and *X. tropicalis Esy1b*, as noted in additional files 1 and 3).

Sequence comparison: multiple alignments

In order to illustrate the conserved features of the proteins in this collection, I used Multalin [57] to construct multiple alignments of subsets of sequences, including expressed variants (additional files 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 and 50). The conserved groups of vertebrates and invertebrates are compared separately but the *Syt18* group (additional file 37) contains members from vertebrates and invertebrates. All of the *Syt* sequences from *N. vectensis* (additional file 41) are compared. These multiple alignments illustrate the patterns of sequence conservation (indicating the action of selection) distinctive of different groups. Common patterns of alternative splicing which serve to alter particular domains, are also illustrated. Where calcium coordinating residues are absolutely conserved (indicating the action of purifying selection on calcium binding activity) these are indicated by arrows. These absolutely conserved sequences are highlighted in red in figures 1, 2 and 3. In the many more cases where calcium coordinating positions are not absolutely conserved, there is a range, down to zero conservation, of the ten sites present in the absolutely conserved, calcium responsive *Syt* proteins. A conserved region 20-50 residues upstream of the C2A domain, often recognisable as a variant of this motif: GRIKPELY is indicated where present. Conservation of this motif was noticed previously [30] but it is obvious here, that this region is much more widely conserved. Its function is yet to be investigated. The position and phase of introns are also illustrated, in order to show the gene structures distinctive of different family groups, as well as instances of intron mobilization. Intron mobilization within a genus is specifically noted

in additional file 2. Intron phase is indicated by colour: black for phase 0, red for phase +1 and blue for phase +2.

Syt1 homologues

Additional file 9 shows a multiple alignment of the invertebrate *Syt1* protein sequences in this collection. The ten conserved calcium coordinating positions are indicated by arrows. The C2AB region is highly conserved whereas the N-terminal portion has much more sequence variation. Expressed variants are included to demonstrate the regions altered by alternative splicing and RNA editing. The region between the TM domain and the C2A domain is commonly altered, for instance in *S. purpuratus*, *B. floridae*, *A. gambiae* and *D. melanogaster*. The next region expressed as alternative forms, is the region between C2A and C2B, as indicated by *D. pulex* and *A. gambiae* variants. The third region to be altered, by alternative splicing or by RNA editing, is the central portion of the C2B domain. In *Caenorhabditis*, this recoding alters not only the identity of positions within this exon of 45 amino acids length in all other cases, but extends its length by 2 or 3 amino acids. Additional file 9 also includes the *snt-3* sequences from *Caenorhabditis*. The *snt-3* gene encodes a protein very similar to the *snt-1* protein which is the *Syt1* orthologue. The *snt-3* proteins however, lack the N-terminal membrane anchoring portion. The *snt-3* proteins also differ in the region between C2A and C2B.

Additional file 10 shows a multiple alignment of the vertebrate *Syt1* protein sequences in this collection. The ten conserved calcium coordinating positions are indicated by arrows. The C2AB region is highly conserved whereas the N-terminal portion has a little more sequence variation. Expressed variants are included to demonstrate the regions altered by alternative splicing. This time, the alteration is restricted to the region between the TM and C2A domain. Conserved N-glycosylation consensus sites are indicated by blue boxes. In the first coding exon, the N-glycosylation site and the conserved upstream O-glycosylated threonine, are known to be important functional sites *in vivo* [25,71,72].

Additional files 11, 12 and 13 show multiple alignments of the vertebrate *Syt2*, *Syt5* and *Syt8* protein sequences, respectively. For *Syt2*, *Syt5* and *Syt8* genes, there is transcript evidence of alternatively expressed forms which lack the TM domain. *Syt1*, *Syt2*, *Syt5* and *Syt8* are all related by duplication, with *Syt1* being the parental gene (figure 4). In the fish genomes, there are yet more duplicates. In order to distinguish the relationships among these duplicates, pairwise comparison scores (figure 5) and syntenic relationships (figure 6) were examined. The top scores highlighted in blue in figure 5, indicate that *D. rerio syt1a* is orthologous to

Human gene	Linked to SY77 in <i>H.sapiens</i> ?	NCBI accession	Linked to syt11 in <i>X.tropicalis</i> ?	JGI gene model	Homologue in <i>D.erio</i> ?	NCBI accession	Linked to which syt in <i>D.erio</i> ?
<i>SY71</i>	Yes - chr12, 78,106K	NM_005639.1	Yes - JGI scaff 160:868560-953624	MOL_fgenesh1_pg.C_scaffold_160000017	Yes - two		<i>syt1a/syt1b</i>
<i>PPP1R12A</i>	Yes - chr12, 78,672K	NM_002480.1	Yes - JGI scaff 160:3475-57288	e_gw1.611.23.1	Yes - chr4, 29,115K	NM_001003870.1	<i>syt1a</i> - chr4, 29,382K
<i>PAWR</i>	Yes - chr12, 78,497K	NM_002583.2	Yes - JGI scaff 160:1007608-1039219	e_gw1.160.91.1	Yes - chr4, 29,226K	NM_001006015.1	<i>syt1a</i> - chr4, 29,382K
<i>NAV3</i>	Yes - chr12, 76,700K	NM_014903.4	Yes - JGI scaff 160:530993-568040	e_gw1.160.72.1	Yes - chr4, 29,701K	NM_001045143.2	<i>syt1a</i> - chr4, 29,382K
<i>BBS10</i>	Yes - chr12, 75,261K	NM_024685.3	Yes - JGI scaff 541:703468-708035	estExt_Genewise1_C_5410005	Yes - chr18, 5,679K	NM_001089463.1	<i>syt1b</i> - chr18, 5,667K
Human gene	Linked to SY72 in <i>H.sapiens</i> ?		Linked to syt2 in <i>X.tropicalis</i> ?		Homologue in <i>D.erio</i> ?		Linked to which syt in <i>D.erio</i> ?
<i>SY72</i>	Yes - chr1, 200,817K	NM_177402.4	Yes - JGI scaff 18:5027186-5045705	MOL_e_gw1.18.132.1	Yes - two		<i>syt2a/syt2b</i>
<i>PPP1R12B</i>	Yes - chr1, 200,554K	NM_002481.2	Yes - JGI scaff 18:4973914-5008317	fgenesh1_kg.C_scaffold_180000015	Yes - chr6, 46,447K	NM_001077579.1	<i>syt2b</i> - chr6, 37,840K
<i>TNN1</i>	Yes - chr1, 199,637K	NM_003281.3	Yes - JGI scaff 18:5255634-5264033	e_gw1.18.234.1	Yes - chr6, 38,208K Also - chr23, 4,292K	chr6: NM_001008613.1 chr23: XM_681611.2	<i>syt2a</i> and <i>syt2b</i>
<i>TNN2</i>	Yes - chr1, 199,592K	NM_000364.2	Yes - JGI scaff 18:5283990-5308657	fgenesh1_pg.C_scaffold_18000127	Yes - chr23, 39,003K	NM_152893.1	<i>syt2a</i> - chr23 but not close, 1,917
<i>PKP1</i>	Yes - chr1, 199,513K	NM_000299.2	Yes - JGI scaff 18:5317831-5332694	fgenesh1_pg.C_scaffold_18000128	Yes - chr6, 38,274K	XM_001338263.2	<i>syt2b</i> - chr6, 37,840K
<i>NAV1</i>	Yes - chr1, 199,862K	NM_020443.3	Yes - JGI scaff 18:5107431-5177840	e_gw1.18.116.1	Yes - chr6, 38,002K Also - chr23, 33,056K	chr6: XM_001920952.1 chr23: XR_044873.1	<i>syt2a</i> and <i>syt2b</i>
Human gene	Linked to SY75 in <i>H.sapiens</i> ?		Linked to syt5 in <i>X.tropicalis</i> ?		Homologue in <i>D.erio</i> ?		Linked to which syt in <i>D.erio</i> ?
<i>SY75</i>	Yes - chr19, 60,375K	NM_003180.2	Yes - JGI scaff 356:560155-582080	MOL_e_gw1.356.4.1	Yes - two		<i>syt5a/syt5b</i>
<i>PPP1R12C</i>	Yes - chr19, 60,290K	NM_017607.2	Yes - JGI scaff 356:419431-424397	e_gw1.356.128.1	Not in assembly Zv7		
<i>TNN1</i>	Yes - chr19, 60,334K	NM_003283.4	Yes - JGI scaff 356:462258-467771	e_gw1.356.16.1	Yes - chr3, 29,898K	NM_001128695.1	<i>syt5a</i> - chr3, 29,884K
<i>TNN3</i>	Yes - chr19, 60,354K	NM_000363.4	Yes - JGI scaff 356:508771-523606	fgenesh1_pg.C_scaffold_356000024	Not in assembly Zv7		
<i>PTRH</i>	Yes - chr19, 60,380K	NM_002842.3	Not in assembly v4.1		Yes - chr24, two copies, 6,753K and 6,665K	XM_001919198.1	<i>syt5b</i> - chr24, 6,788K
Human gene	Linked to SY78 in <i>H.sapiens</i> ?		Linked to syt8 in <i>X.tropicalis</i> ?		Homologue in <i>D.erio</i> ?		Linked to which syt in <i>D.erio</i> ?
<i>SY78</i>	Yes - chr11, 1,811K	NM_138567.3	Yes - JGI scaff 419:462072-469381	MOL_estExt_fgenesh1_pm.C_4190008	Yes - chr25, 26,331K		<i>syt8</i>
<i>TNN3</i>	Yes - chr11, 1,895K	NM_006757.3	Yes - JGI scaff 419:562392-571544	e_gw1.419.79.1	Yes - chr25, 26,831K	NM_131565	<i>syt8</i>
<i>TNN2</i>	Yes - chr11, 1,817K	NM_003282.3	Yes - JGI scaff 419:473297-481004	estExt_fgenesh1_pg.C_4190018	Yes - chr25, 26,659K	NM_001007365.1	<i>syt8</i>

Figure 6 Synteny of *Syt1* paralogues in *D. rerio*, *X. tropicalis* and *H. sapiens*. Gene symbols for four groups of neighbouring genes in *H. sapiens*, are enclosed by a black box. Within the red box, the chromosomal locations of each gene in the current human genome reference sequence, are indicated along with a reference transcript sequence. The genomic locations and transcript sequences for the *X. tropicalis* gene relatives are within the green box and those for *D. rerio* are within the blue box.

While there is some variation in length, of sequences from different organisms in the C2B region, the more pronounced difference is a trend towards increased length of the region between TM and C2A domains in insects. Additional file 25 shows the vertebrate *Syt12* sequences. Transcript evidence indicates the expression of forms varying in the N-terminal region in *H. sapiens*. Little functional work has been done so far with *Syt12* [19,82].

Syt13 homologues

Additional file 26 shows a multiple alignment of the invertebrate *Syt13* protein sequences in this collection. In most of these genes, there are two alternative N-terminal coding exons, only one of which encodes a TM domain. These sequences also retain the motif of unknown function just upstream of the C2A domain. Additional file 27 shows the vertebrate *Syt13* sequences. Transcript sequences from *H. sapiens* indicate the expression of a form lacking the TM domain. A recent study suggests that human *SYT13* may function as a tumour suppressor [83].

Syt15 homologues

Additional file 28 shows a multiple alignment of the invertebrate *Syt15* protein sequences in this collection. In these sequences, it is the C2A domain which is more highly conserved than the C2B domain. The motif upstream of the C2A domain is present. There is a large amount of variation in amino acid sequence, intron position and intron phase in the N-terminal portion of these sequences. Additional file 29 shows the vertebrate *Syt15* sequences. Transcript evidence from *M. musculus* and *H. sapiens* indicates the expression of variants

differing at the C-terminal end. Transcripts from *H. sapiens* indicate variant forms at the N-terminal end, which lack the TM domain. Very little is known about the function of *Syt15* [19].

Syt16 homologues

Additional file 30 shows a multiple alignment of the invertebrate *Syt16* protein sequences in this collection. Transcript evidence indicates the expression of variants in *C. savignyi*, *C. intestinalis* and *S. purpuratus* which differ at the N-terminal end and lack the TM domain. Additional file 31 shows the vertebrate *Syt16* sequences which contain a TM domain. Additional file 32 shows the vertebrate *Syt16* sequences which lack a TM domain. Additional file 33 shows vertebrate *Syt14* sequences. Transcript evidence in *X. tropicalis*, *M. musculus* and *H. sapiens* indicates the expression of variant forms of *Syt14* with altered C2B domains. Human *SYT14* transcripts also differ at the N-terminal region, resulting in forms which lack the TM domain. Loss of function of human *SYT14* has been associated with neurodevelopmental abnormalities [84]. In the human genome, there is a repeat of a portion of the *SYT14* gene (which is on chromosome 1) on chromosome 4, apparently the result of retroposition of a *SYT14* transcript. The chromosome 4 locus has been named *SYT14L* by HGNC [31]. *SYT14L* is classed as a pseudogene by NCBI (GeneID: NR_027094) and is not included in this collection.

Syt17 homologues

Additional file 34 shows a multiple alignment of the invertebrate *Syt17* protein sequences in this collection. A possible motif upstream of the C2A domain is

indicated. Additional file 35 shows the vertebrate Syt17 sequences. Transcript evidence in *M. musculus* and *H. sapiens* indicates the expression of variants altered at N-terminal and C-terminal ends. A possible motif upstream of the C2A domain is indicated. The protein product of Syt17 has been referred to as B/K protein, but not much is known about its function [85].

Sytalpha homologues

Additional file 36 shows the Sytalpha sequences in this collection. A possible motif upstream of the C2A domain is indicated. The only functional study relevant to Sytalpha is its anatomical localisation in *D. melanogaster* [27].

Syt18 homologues

Additional file 37 shows the Syt18 sequences in this collection. These proteins are so far completely uncharacterised.

Syt19 homologues

Additional file 38 shows the vertebrate Syt19 sequences in this collection. These proteins are so far completely uncharacterised.

Syt21 homologues

Additional file 39 shows the invertebrate Syt21 sequences in this collection. These proteins are so far completely uncharacterised.

Syt homologues in *N. vectensis*

Additional file 40 shows all of the Syt proteins of *N. vectensis*, demonstrating enormous sequence variation. These proteins are so far completely uncharacterised.

Dblc2 homologues

Additional file 41 shows the Dblc2 sequences in this collection. Transcript variants are indicated. These proteins have not previously been reported and are so far completely uncharacterised. Dblc2 sequences are detectably similar to Syt sequences at the amino acid sequence level and occur in the genomes of marine invertebrates in this collection. While they lack an N-terminal TM domain, they have a tandem C2 domain architecture in common with Syt proteins. *Dblc2* genes lack similarity to *Syt* genes at the gene structure level, so a new gene symbol was chosen to represent these genes. The Dblc2 designation represents a fuller description: double C2 domain protein. Dblc2 is thus similar but recognisably different from Doc2, the designation for the homologous family of double C2 domain proteins of vertebrates.

Esyt homologues

Additional file 42 shows the invertebrate Esyt2 sequences in this collection. Transcript variants are indicated. The arthropod *Esyt2* genes include mutually exclusive alternative exons which serve to alter the specification of the first C2 domain. Additional file 42 shows the sequence of the first of these alternatives only. Additional file 43 shows all versions of this

alternatively coded region. Additional file 44 shows the vertebrate Esyt1 sequences. Transcript variants are indicated. In the fish genomes, *Esyt1* has duplicated, resulting in *esyt1a* and *esyt1b*. The *esyt1b* duplicate has acquired an internal duplication which contributes an additional four C2 domains to the product of this gene [33]. Within this duplicated section, the pink dot marks the position of an apparent phase 0 intron loss in the tetraodontiform fish. In each of the three other occurrences (two prior, one subsequent) of this portion of the *esyt1b* gene, the phase 0 intron is present. This intron is also present in each of the two repeats of this sequence, in the *esyt1a* genes. Additional file 45 shows the vertebrate Esyt2 sequences. Transcript variants are indicated. In the fish genomes, *Esyt2* has duplicated, resulting in *esyt2a* and *esyt2b*. Additional file 46 shows the vertebrate Esyt3 sequences. Transcript variants are indicated.

Similarity between *Syt* genes and what are now named *Esyt* genes, was first evident from genome sequence comparisons [29,30]. In trying to classify and annotate the genes within the human genome, HGNC [31] noticed that these genes formed a distinct gene family, separate from *Syt* genes. As there was little functional data associated with these genes, a temporary nomenclature with no functional implication was provided: *FAM62A*, *FAM62B* and *FAM62C*. The functional data which did exist, was in the form of a cDNA representing Rat *Esyt1*, which had been cloned during an investigation of adipocyte proteins and found to encode a membrane bound C2 domain protein with similarity to proteins in plants [86]. Further genome sequence comparisons indicated that *Syt*, *FAM62* and plant *NTMC2* genes encode proteins with amino acid sequence similarity and domain architecture similarity, but belong to distinct gene families [33]. Apart from [86] and until recently, functional data associated with plant *NTMC2* and animal *FAM62* gene families had been lacking. An initial investigation of proteins encoded by the Human *FAM62A*, *FAM62B* and *FAM62C* genes, has now been published and the authors named these proteins Esyt1, Esyt2 and Esyt3, to stand for extended synaptotagmin-like proteins [67]. Consultation with the nomenclature committees during the preparation of this paper, led to the decision to implement the *Esyt* nomenclature in place of *Fam62*. In an initial pair of studies on one member of the plant *NTMC2* gene family [34,35] the authors put forward the name SYT1 for this particular gene and gene product. As discussed in the background section of this paper, when different communities make naming decisions based on the criteria most relevant to them (gene phylogeny, protein structure, biological function) it is inevitable that multiple names will be used to identify the same entities.

Rph3a homologues

Additional file 47 shows the invertebrate Rabphilin sequences in this collection. In *Caenorhabditis*, alternative N-terminal exons can be used. Sequence conservation among Rabphilins is high in the N-terminal, Rab binding portion [87] and the C-terminal tandem C2 domain portion, but a large middle portion is very poorly conserved. This means that where transcript confirmation is absent, the gene predictions across this portion are unlikely to be accurate. Additional file 48 shows the vertebrate Rph3a sequences. Transcript variants are indicated. In vertebrates, all 10 calcium coordinating positions are absolutely conserved, but not all are conserved in invertebrates. In additional file 47, the positions of these 10 amino acids are indicated by pink dots. The position of the fifth calcium coordinating residue in the C2A domain is not strictly conserved, but in most cases a suitable aspartate or glutamate residue is present one residue earlier. I am not aware of biochemical evidence for calcium binding by invertebrate Rabphilin proteins, but genetic evidence [88] suggests that *C. elegans rbf-1* at least, appears to function similarly to its mammalian homologues. For this reason, in figures 1, and 2, I have counted the fifth position as present when a suitable aspartate or glutamate residue occurs at the usual spacing or one residue earlier.

Rph3al homologues

Additional file 49 shows vertebrate Rph3al sequences. Transcript variants are indicated. In the functional literature, products of *Rph3al* genes have been referred to as Noc2 [89,90]. As outlined above, *Rph3al* genes appear to be the result of duplication and fission of an ancestral *Rph3a* gene, producing linked *Rph3al* and *Doc2b* genes in tetrapods and in *D. rerio*. *Rph3al* represents the N-terminal portion of the ancestral *Rph3a* gene and *Doc2b* represents the C-terminal portion, as seems clear from a comparison of the gene structures of *Rph3a*, *Rph3al* and the linked *Doc2b* genes (additional files 48, 49, and 50). As also indicated above, *Doc2b* genes do not always accompany *Rph3al* genes, as they exist in other fish genomes where *Rph3al* genes do not occur (see figure 3).

Doc2 homologues

Additional file 50 shows the vertebrate Doc2 protein sequences. Transcript variants are indicated as well as the positions of the calcium coordinating residues, which are conserved in all of the Doc2 proteins except Doc2g. I have included potential Doc2g products from the human gene locus. Although spliced transcripts are produced from this locus, they cannot encode full length proteins and it is not known whether protein products are translated from these transcripts. It is clear from the shared gene structure of *Doc2* genes and *Rph3a* genes, that *Doc2* genes represent the C-terminal portion of a *Rph3a* gene. Since *Rph* genes exist in invertebrates, but

Doc2 genes do not, it seems reasonable to suggest that *Doc2* genes arose via duplication and fission of a *Rph3a* gene. Evidence for this, in the form of the two linked fission products (*Rph3al* and *Doc2b*) is present in the genomes of *D. rerio*, *X. tropicalis*, *A. carolinensis*, *G. gallus*, *T. guttata*, *O. anatinus*, *M. domestica*, *M. musculus* and *H. sapiens*. In this collection, the fish genomes reveal a novel lineage of *Doc2* genes, which in consultation with Zfin staff [62,63] were given the gene symbol *Doc2d*, to stand for Doc2 delta. The first *Doc2* gene, cloned in 1995 [91] was soon found to represent the first member of a family of three genes in rodents. The protein products of these three genes were named Doc2 alpha, Doc2 beta and Doc2 gamma [92]. The conserved region present in the N-terminal portion of *Doc2* sequences, specifies the Munc13 binding capacity of *Doc2* proteins [93].

Conclusions

This paper describes a manually curated collection of genes pertinent to intercellular trafficking in multicellular animals. The collection includes 141 genes from the genome sequences of 9 marine invertebrates (additional file 1) 215 genes from the genome sequences of 24 ecdysozoans, including the 12 *Drosophila* and 5 *Caenorhabditis* genomes (additional file 2) plus 355 genes from 13 vertebrate genomes (additional file 3). When viewed overall (figures 1, 2 and 3) the collection implies that particular homologues which are always present in the genome repertoire, serve to specify functions required for animal life. The basis for this view is that these genes have not been lost, but have been retained and conserved, over the long period of metazoan evolution. Such a view about the essentiality of gene function, differs from a synthetic biology view, which considers the genes necessary to specify an operable system which functions outside the natural world. It differs too, from a reductionist molecular biology view where, if a gene can be deleted in the laboratory setting, without killing the organism, it is not considered essential. Such gene deletion experiments have been done with *Syt1* in different organisms [94-96] and according this view, *Syt1* is not strictly essential to animal life because *snt-1* null mutants are not lethal in *Caenorhabditis*. In *Drosophila* and *Mus* however, *Syt1* null mutants are indeed lethal as they cannot survive to reproduce as adults. While my previous genomic comparison of *Syt* genes [30] could have been interpreted as indicating a correlation between increased numbers of *Syt* genes and increased organism complexity [33] the analysis here, which is much more comprehensive, definitively rules that conclusion out. This collection of genes offers prospects for fruitful speculation about the functional roles of the *Syt* repertoires of different animals and includes a great

range of biological complexity. The conclusions drawn from evolutionary genomics approaches are always provisional, because they depend on the quality of information available (draft or 'complete' genome sequence, availability of transcript information) and are thus subject to review when better information becomes available. Conclusions drawn from reassembled and reannotated genomes will inevitably be more accurate than those drawn from the first draft. The *D. rerio* genome being repeat rich, is taking time to assemble into a complete version. Among the genomes in this collection, the draft *D. rerio* genome ranks first in terms of gene number. *D. rerio* surely presents a useful model system with which to investigate the functional divergence of close paralogues [41].

Currently, it is the rodent Syt proteins and, with the advent of transgenic mouse studies, their respective genes, which garner most interest and have been best characterised. Work with *Syt* genes in other animal species has provided information supporting hypotheses about how *Syt* genes fulfil their functional roles [94,95,97,98]. To date, these studies indicate a primary role for *Syt1* in specifying a neuron specific synaptic vesicle protein which serves to trigger exocytosis. The genomic specification of *Syt1* expression patterns, has been investigated and utilised in ascidian species [99-102]. In *C. elegans*, the anatomical expression patterns of the mutually exclusive alternative exon variants of *Syt1* have been mapped [103]. This gene collection shows that such mutually exclusive exon deployment has arisen on a number of occasions in invertebrates, altering C2 domains in *Syt1*, *Syt34* and *Esyt2* proteins (see additional file 5). The repeated evolution of this type of C2 domain alteration, suggests that careful comparison of the altered forms might prove a useful focus for investigations aimed at understanding how these C2 domains operate as molecular machines. The feature most widely shared among the great diversity of Syt proteins in this collection, is sequence conservation of the C-terminal C2B region, which suggests that a defining function of Syt proteins is located here.

The regulatory machinery controlling variant expression can be illuminated by comparative analyses of genomic sequences, as was done to investigate the RNA editing of *Syt1* which occurs in hexapods [104]. Another comparative analysis [105] examining the genomic specification of alternative splicing of *Syt1*, concluded that splice variants expressed in *D. melanogaster*, *C. pipiens*, *B. mori*, *T. castaneum* and *P. humanus* depend on correctly coded intronic, *cis*-acting regulatory sequences acting on the requisite splice donor/acceptor sequences. In this collection, there is evidence from *A. gambiae*, *D. melanogaster*, *D. pulex*, *S. purpuratus*, *B. floridae*, *D. rerio*, *X. tropicalis*, *A. carolinensis*, *G. gallus*, *M.*

musculus and *H. sapiens* of RNA editing and alternative splicing of *Syt1* (additional files 1, 2, 3, 9 and 10). These sequences may aid future efforts to identify the *cis*-sequences involved in transcript processing. The biochemical understanding of transcript production and processing, acquired at great pains over many years [see [43]] is not quite up to the task of extracting the full meaning from the flood of new transcript data from high throughput sequencing. An enormous amount of experimental work will need to be done to understand the varied roles of transcripts in complex organisms. The problem of identifying and annotating different kinds of transcript sequences is discussed in [106]. Work to assess the functional significance of alternatively expressed *Syt* genes has barely begun [103,107-112] so it is difficult at present, to judge the importance of the transcript variants listed in this collection. Evolutionary conservation of variant production, probably argues for functional significance. New developments in mass spectrometry [113] are bound to aid attempts to examine complex mixtures of proteins such as synaptic vesicle preparations [114,115]. It is more likely however, that the functional impact of Syt protein variants will eventually be revealed by research focussed on Syt proteins of known importance.

The significance of the conservation of intron position and phase within *Syt* genes could be investigated by gene targeting in transgenic mice, now a common route to investigate *Syt* roles [6,21]. The best conserved intron, a signature feature of *Syt* genes, is the phase +1 intron at the start of the third beta strand of the C2B domain [29,33,36] (additional files 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 and 41). A single point mutation to inactivate this intron acceptor dinucleotide would be predicted to have devastating consequences for the expression of the host gene. In contrast to the approach of disabling a gene by creating a large deletion, the production of a null mutant by such a small intervention would be unlikely to cause collateral damage. The chances of cryptic splice acceptors being deployed to restore normal function to such a targeted *Syt* splice site point mutant, seem remote, based on the importance of this region to Syt protein function. The presence of this intron in so many *Syt* genes could be interpreted as evidence that evolutionary alteration of this site is hazardous.

An example of a gene inactivating mutation which is not rescued by altered splicing, is present in this gene collection. The *H. sapiens* gene which is orthologous to *M. musculus Doc2g* has acquired a frameshift mutation in the fifth coding exon which introduces a stop codon. Human transcript sequences indicate that an alternative intron acceptor dinucleotide capable of restoring the

reading frame, which exists 20 nucleotides upstream of the exon 5 acceptor, is not used. The regulatory sequences necessary to select this alternative acceptor are not present within this short intron (figure 7). It is not known what protein products, if any, are produced by this gene in *H. sapiens* and it appears likely that the human gene is non functional. This observation underscores the importance of splicing signals which are poorly understood at present, as well as the importance of transcript analysis to verify gene predictions.

In sea urchin species, the *Syt1* protein has been shown to act in cortical granule exocytosis [98]. *Syt1* products are used as pan neuronal markers in echinoderms, hemichordates and urochordates [25] and the role of *Syt1* in synaptic vesicle exocytosis is assumed. Few comparative studies have attempted to map the expression patterns of the *Syt* repertoire of a particular organism but all of the *Syt* genes of *D. melanogaster* [27] and a large fraction of those of *R. norvegicus* [28] have been examined in this way. *T. adhaerens* is thought to lack cells equivalent to neurons, but it has a well conserved *Syt1* as well as genes encoding the requisite neuronal SNAREs and complexin for fast neurotransmission (additional file 4). Although never observed in the laboratory, *T. adhaerens* is expected to produce gametes in the wild [116,117] so a role for the *T. adhaerens Syt1* similar to that of sea urchin *Syt1* in cortical granule exocytosis is possible, alongside any yet to be discovered 'neural' role. *N. vectensis*, which has a bona fide nervous

system, also has *Syt1* and the requisite neuronal SNAREs and complexin (additional file 4).

N. vectensis is quite unusual among the metazoans collected here, in having so many highly divergent *Syt* genes (additional file 41). Of all the *N. vectensis Syt* genes, only *Syt1* and *Syt7* encode all ten calcium coordinating positions, indicating their likely role in calcium triggered membrane fusion. Future studies to determine the anatomical expression patterns of the *N. vectensis Syt* genes *in vivo*, will be of great interest to further understand the evolutionary development of animal nervous systems. Initial study of the molecular anatomy of the *N. vectensis* nervous system has examined the expression patterns of neurotransmitters, neuropeptides, transcription factor and cell adhesion genes [118]. Synaptotagmin gene expression was used to label differentiated neurons in *Platynereis*, in a study of the evolutionary development of the central nervous system of animals [119]. *Capitella*, like *Platynereis*, is an annelid member of the Lophotrochozoa, but has the advantage as a model organism, of a whole genome sequence [120]. Study of neurogenesis in *Capitella* is now underway [121]. *In situ* hybridization analysis of *Syt1* expression in *Capitella*, shows a broad neural pattern (Elaine Seaver, personal communication).

It will be important in future, to describe and compare anatomical expression patterns and functions of *Syt* genes, both conserved and unique, in vertebrates and invertebrates. Such comparisons will be essential

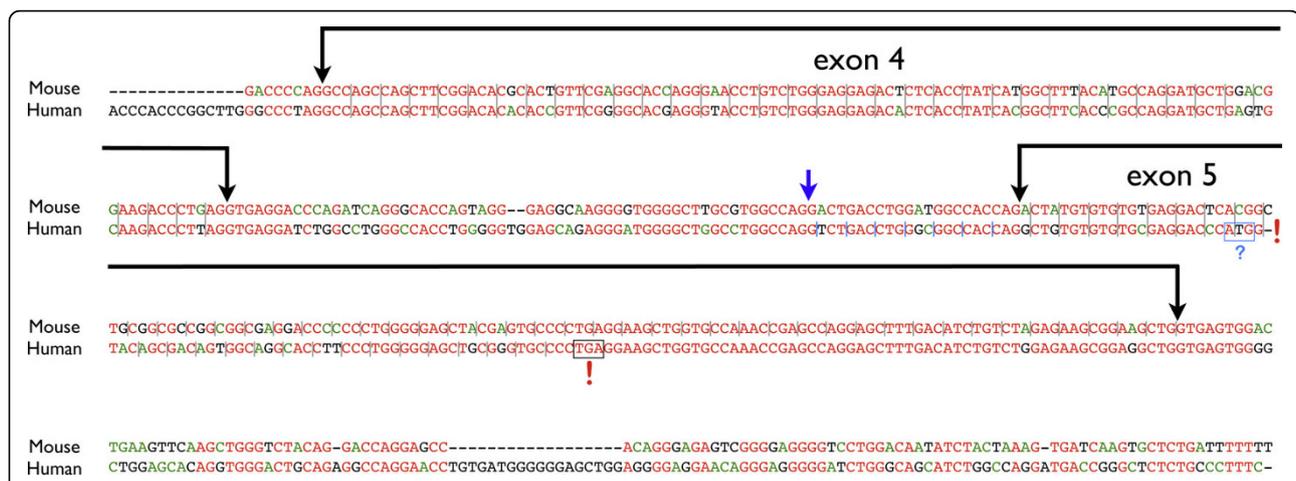


Figure 7 Genomic sequence surrounding coding exons 4 and 5 in mouse *Doc2g* and the equivalent human region. Genomic sequence surrounding coding exons 4 and 5 of mouse *Doc2g* and the equivalent region in human, are shown. Exonic and intronic sequences are indicated. The reading frame is indicated by grey bars between codons. The single nucleotide deletion in the human exon 5 region is highlighted with a red exclamation mark. The deletion in human exon 5 disrupts the reading frame, leading to a premature termination codon which is boxed and highlighted with a red exclamation mark. An alternative splice acceptor dinucleotide which could restore the correct reading frame to human exon 5 is indicated by a blue arrow. Blue bars between the first few codons indicate the reading frame were the alternative splice acceptor to be functional. Human exon 5 has an alternative start codon, which is boxed and highlighted with a question mark. There is a possibility that this start codon could allow normally spliced transcripts to be translated into the N-terminal and C-terminal protein products listed in additional file 50.

to improve, or reduce confidence in functional annotation of genomes based on phylogenetic homology, or similarity otherwise defined. Good annotation of genomes is a necessary step in the process of biocuration [122] a new field aimed at using twenty first century sequencing and information technology to make accessible to bench scientists, the vast amount of information they require, distilled, in order to take forward wisely informed hypothesis driven research. Community deposition of a variety of types of biological data into appropriate databases, database efforts to curate, update and integrate this data with relevant data in other databases, will if pursued successfully, lead to great progress in biology in the years ahead. For the time being, the manual gene collection made available here, can be used immediately, by bench scientists engaged in hypothesis driven investigation of membrane trafficking by these proteins, to widen the bounds of a field already stretching from biophysical study *in vitro* at the atomic level, to phenotypic assessment of gene function in model organisms.

Additional file 1: Comprehensive information about the marine invertebrate genes in this collection. This spreadsheet file (MarineInvertebrates.xls) contains full details of each gene identified in the marine invertebrate genomes in this collection.
Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S1.XLS>]

Additional file 2: Comprehensive information about the ecdysozoan genes in this collection. This spreadsheet file (Ecdysozoa.xls) contains full details of each gene identified in the ecdysozoan genomes in this collection.
Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S2.XLS>]

Additional file 3: Comprehensive information about the vertebrate genes in this collection. This spreadsheet file (Vertebrates.xls) contains full details of each gene identified in the vertebrate genomes in this collection.
Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S3.XLS>]

Additional file 4: Comprehensive information about the SNARE and complexin genes identified in *T. adhaerens* and *N. vectensis*. This spreadsheet file (SNAREs.xls) contains full details of the genes identified in *T. adhaerens* and *N. vectensis* which are similar to neuronal SNAREs and complexin.
Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S4.XLS>]

Additional file 5: Dendrogram of relationships among the invertebrate sequences in this collection. Additional file 5 displays the guide tree of the clustalw2 comparison of the invertebrate sequences in this collection, excluding variants, totalling 356 sequences. Genes which encode mutually exclusive alternative exons are highlighted with a green box. Conserved groups of *Syt* genes which have not previously been described, are highlighted with a blue box.
Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S5.PDF>]

Additional file 6: Dendrogram of relationships among the vertebrate sequences in this collection. Additional file 6 displays the guide tree of the clustalw2 comparison of the vertebrate sequences in this collection, excluding variants, totalling 355 sequences. Mouse genes are highlighted with a red box. Conserved groups of *Syt* genes which have not previously been described, are highlighted with a blue box.
Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S6.PDF>]

Additional file 7: Dendrogram of relationships among the invertebrate and mammalian sequences in this collection. Additional file 7 displays the guide tree of the clustalw2 comparison of the invertebrate sequences (one representative per genus) excluding variants, plus mammalian sequences, excluding variants, totalling 289 sequences. Mouse genes are highlighted with a red box. Conserved groups of *Syt* genes which have not previously been described, are highlighted with a blue box.
Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S7.PDF>]

Additional file 8: Dendrogram of relationships among the invertebrate and vertebrate sequences in this collection. Additional file 8 displays the guide tree of the clustalw2 comparison of the invertebrate sequences (one representative per genus) excluding variants, plus a subset of vertebrate sequences, excluding variants, totalling 454 sequences. Mouse genes are highlighted with a red box. Conserved groups of *Syt* genes which have not previously been described, are highlighted with a blue box.
Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S8.PDF>]

Additional file 9: Alignment of the invertebrate *Syt1* sequences, plus *Caenorhabditis snt-3*. Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. The five conserved acidic amino acids in each C2 domain are indicated by black arrows at the top of the alignment. X residues indicate where a portion of sequence is missing.
Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S9.PDF>]

Additional file 10: Alignment of the vertebrate *Syt1* sequences. Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. The five conserved acidic amino acids in each C2 domain are indicated by black arrows at the top of the alignment. Conserved N-glycosylation consensus sites are indicated by blue boxes. The conserved threonine, which can be O-glycosylated is also indicated by a blue box. X residues indicate where a portion of sequence is missing.
Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S10.PDF>]

Additional file 11: Alignment of the vertebrate *Syt2* sequences. Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. The *X. tropicalis* splice variant is included and highlighted with a black dot where it differs. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. The five conserved acidic amino acids in each C2 domain are indicated by black arrows at the top of the alignment. X residues indicate where a portion of sequence is missing.
Click here for file

approximately, at the top of each page of the alignment. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. The widely conserved motif of unknown function, just upstream of the C2A domain, is indicated. The five conserved acidic amino acids in each C2 domain are indicated by black arrows at the top of the alignment.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S23.PDF>]

Additional file 24: Alignment of the invertebrate Syt12 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. Because of their differing positions, TM domains are highlighted in blue. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S24.PDF>]

Additional file 25: Alignment of the vertebrate Syt12 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. The *H. sapiens* splice variant is included and highlighted with a black dot where it differs. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S25.PDF>]

Additional file 26: Alignment of the invertebrate Syt13 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants which specify an alternative N-terminus lacking a TM domain, are indicated on top, with a dotted line indicating where they join the common C-terminal portion. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. The widely conserved motif of unknown function, upstream of the C2A domain, is indicated. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S26.PDF>]

Additional file 27: Alignment of the vertebrate Syt13 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. The *H. sapiens* splice variant is included and highlighted with a black dot where it differs. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. The widely conserved motif of unknown function, just upstream of the C2A domain, is indicated. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S27.PDF>]

Additional file 28: Alignment of the invertebrate Syt15 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. Because of their differing positions, TM domains are highlighted in blue. The widely conserved motif of unknown function, just upstream of the C2A domain, is indicated. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S28.PDF>]

Additional file 29: Alignment of the vertebrate Syt15 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. The widely conserved motif of unknown function, just upstream of the C2A domain, is indicated. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S29.PDF>]

Additional file 30: Alignment of the invertebrate Syt16 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants which specify alternative N-termini lacking TM domains, are included and highlighted with black dots where their sequences differ. TM domains are highlighted in blue. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S30.PDF>]

Additional file 31: Alignment of the vertebrate Syt16 sequences which have TM domains.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S31.PDF>]

Additional file 32: Alignment of the vertebrate Syt16 sequences which lack TM domains.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S32.PDF>]

Additional file 33: Alignment of the vertebrate Syt14 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S33.PDF>]

Additional file 34: Alignment of the invertebrate Syt17 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. The widely conserved motif of unknown function, upstream of the C2A domain, is indicated. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S34.PDF>]

Additional file 35: Alignment of the vertebrate Syt17 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. A possible motif just upstream of the C2A domain, is indicated. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S35.PDF>]

Additional file 36: Alignment of the invertebrate Sytalpha sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. The widely conserved motif of unknown function, just upstream of the C2A domain, is indicated. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S36.PDF>]

Additional file 37: Alignment of the Syt18 sequences. Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S37.PDF>]

Additional file 38: Alignment of the vertebrate Syt19 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. The X residue indicates where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S38.PDF>]

Additional file 39: Alignment of the invertebrate Syt21 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. TM sequences are highlighted in blue. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S39.PDF>]

Additional file 40: Alignment of the *N. vectensis* Syt sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. TM sequences are highlighted in blue. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S40.PDF>]

Additional file 41: Alignment of the invertebrate Dblc2 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S41.PDF>]

Additional file 42: Alignment of the invertebrate Esyt2 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S42.PDF>]

Additional file 43: Alignment of the mutually exclusive alternative Esyt2 exons.

In addition to the alternatively coded exon which is highlighted, the exon upstream and the exon downstream are shown. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Blue bars indicate phase +2 introns.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S43.PDF>]

Additional file 44: Alignment of the vertebrate Esyt1 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. The middle portion is only present in the fish esyt1b sequences. A pink dot in this portion, marks an intron loss in *T. rubripes* and *T. nigroviridis*. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Blue bars indicate phase +2 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S44.PDF>]

Additional file 45: Alignment of the vertebrate Esyt2 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Blue bars indicate phase +2 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S45.PDF>]

Additional file 46: Alignment of the vertebrate Esyt3 sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. Some intron positions are marked with dotted lines. In these cases, transcript sequence covers a gap in the genomic sequence and the intron presence is assumed. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S46.PDF>]

Additional file 47: Alignment of the invertebrate Rabphilin sequences.

Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are indicated. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. The five conserved acidic amino acids in each C2 domain are indicated by pink dots at the top of the alignment. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S47.PDF>]

Additional file 48: Alignment of the vertebrate Rabphilin (Rph3a) sequences. Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. The five conserved acidic amino acids in each C2 domain are indicated by black arrows at the top of the alignment. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S48.PDF>]

Additional file 49: Alignment of the vertebrate Rph3al sequences. Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S49.PDF>]

Additional file 50: Alignment of the vertebrate Doc2 sequences. Amino acid position is marked every hundred amino acids approximately, at the top of each page of the alignment. Splice variants are included and highlighted with black dots where they differ. Intron position and phase is indicated with a coloured bar between amino acids. Black bars indicate phase 0 introns. Red bars indicate phase +1 introns. Blue bars indicate phase +2 introns. Some intron positions are marked with dotted lines. In these cases, transcript sequence covers a gap in the genomic sequence and the intron presence is assumed. The five conserved acidic amino acids in each C2 domain are indicated by black arrows at the top of the alignment. X residues indicate where a portion of sequence is missing.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-37-S50.PDF>]

Acknowledgements

This work was supported by the UK Medical Research Council. I wish to thank Hugh Pelham for making it possible for me to complete this work.

Authors' contributions

The author had sole responsibility for the design and execution of this work.

Received: 10 August 2009

Accepted: 15 January 2010 Published: 15 January 2010

References

- Matthew WD, Tsavaler L, Reichardt LF: Identification of a synaptic vesicle-specific membrane protein with a wide distribution in neuronal and neurosecretory tissue. *J Cell Biol* 1981, **91**:257-269.
- Südhof TC: The synaptic vesicle cycle. *Annu Rev Neurosci* 2004, **27**:509-547.
- Wojcik SM, Brose N: Regulation of membrane fusion in synaptic excitation-secretion coupling: speed and accuracy matter. *Neuron* 2007, **55**:11-24.
- Chapman ER: How does synaptotagmin trigger neurotransmitter release?. *Annu Rev Biochem* 2008, **77**:615-641.
- Malsam SK, Kreye S, Söllner TH: Membrane fusion: SNAREs and regulation. *Cell Mol Life Sci* 2008, **652**:2814-2832.
- Gustavsson N, Han W: Calcium-sensing beyond neurotransmitters: functions of synaptotagmins in neuroendocrine and endocrine secretion. *Biosci Rep* 2009, **29**:245-259.
- Perin MS, Fried VA, Mignery GA, Jahn R, Südhof TC: Phospholipid binding by a synaptic vesicle protein homologous to the regulatory region of protein kinase C. *Nature* 1990, **345**:260-263.
- Sutton RB, Davletov BA, Berghuis AM, Südhof TC, Sprang SR: Structure of the first C2 domain of synaptotagmin I: a novel Ca²⁺/phospholipid-binding fold. *Cell* 1995, **80**:929-938.
- Sutton RB, Ernst JA, Brunger AT: Crystal structure of the cytosolic C2A-C2B domains of synaptotagmin III: implications for Ca²⁺-independent SNARE complex interaction. *J Cell Biol* 1999, **147**:589-598.
- Nalefski EA, Falke JJ: The C2 domain calcium-binding motif: structural and functional diversity. *Protein Sci* 1996, **5**:2375-2390.
- Jiménez JL, Smith GR, Contreras-Moreira B, Sgouros JG, Meunier FA, Bates PA, Schiavo G: Functional recycling of C2 domains throughout evolution: a comparative study of synaptotagmin, protein kinase C and phospholipid C by sequence, structural and modelling approaches. *J Mol Biol* 2003, **333**:621-639.
- Cho W, Stahelin RV: Membrane binding and subcellular targeting of C2 domains. *Biochim Biophys Acta* 2006, **1761**:838-849.
- Pfam database. <http://pfam.sanger.ac.uk/family/C2>.
- Fuson KL, Montes M, Robert JJ, Sutton RB: Structure of human synaptotagmin 1 C2AB in the absence of Ca²⁺ reveals a novel domain association. *Biochemistry* 2007, **46**:13041-13048.
- Chicka MC, Hui E, Liu H, Chapman ER: Synaptotagmin arrests the SNARE complex before triggering fast, efficient membrane fusion in response to Ca²⁺. *Nat Struct Mol Biol* 2008, **15**:827-835.
- Paddock BE, Streigel AR, Hui E, Chapman ER, Reist NE: Ca²⁺-dependent, phospholipid-binding residues of synaptotagmin are critical for excitation-secretion coupling *in vivo*. *J Neurosci* 2008, **28**:7458-7466.
- von Poser C, Ichtchenko K, Shao X, Rizo J, Südhof TC: The evolutionary pressure to inactivate. A subclass of synaptotagmins with an amino acid substitution that abolishes Ca²⁺ binding. *J Biol Chem* 1997, **272**:14314-14319.
- Rickman C, Craxton M, Osborne S, Davletov B: Comparative analysis of tandem C2 domains from the mammalian synaptotagmin family. *Biochem J* 2004, **378**:681-686.
- Bhalla A, Chicka MC, Chapman ER: Analysis of the synaptotagmin family during reconstituted membrane fusion: uncovering a class of inhibitory isoforms. *J Biol Chem* 2008, **283**:21799-21807.
- Zhang Z, Bhalla A, Dean C, Chapman ER, Jackson MB: Synaptotagmin IV: a multifunctional regulator of peptidergic nerve terminals. *Nat Neurosci* 2009, **12**:163-171.
- Dean C, Liu H, Dunning F, Chang PY, Jackson MB, Chapman ER: Synaptotagmin-IV modulates synaptic function and long-term potentiation by regulating BDNF release. *Nat Neurosci* 2009, **12**:767-776.
- Littleton JT, Bellen HJ, Perin MS: Expression of synaptotagmin in *Drosophila* reveals transport and localization of synaptic vesicles to the synapse. *Development* 1993, **118**:1077-1088.
- Chowdhury D, Travis GH, Sutcliffe JG, Burton FH: Synaptotagmin I and 1B4 are identical: implications for Synaptotagmin distribution in the primate brain. *Neurosci Lett* 1995, **190**:9-12.
- Craxton M, Goedert MG: Alternative splicing of synaptotagmins involving transmembrane exon skipping. *FEBS Lett* 1999, **460**:417-422.
- Burke RD, Osborne L, Wang D, Murabe N, Yaguchi S, Nakajima Y: Neuron-specific expression of a synaptotagmin gene in the sea urchin *Strongylocentrotus purpuratus*. *J Comp Neurol* 2006, **496**:244-251.
- Cnops L, Hu TT, Broeck Vanden J, Burnat K, Bergh Van Den G, Arckens L: Age- and experience-dependent expression of dynamin I and synaptotagmin I in cat visual system. *J Comp Neurol* 2007, **504**:254-264.
- Adolfson B, Saraswati S, Yoshihara M, Littleton JT: Synaptotagmins are trafficked to distinct subcellular domains including the postsynaptic compartment. *J Cell Biol* 2004, **166**:249-260.
- Mittelsteadt T, Seifert G, Álvarez-Barón E, Steinhäuser C, Becker A, Schoch S: Differential mRNA expression patterns of the synaptotagmin gene family in the rodent brain. *J Comp Neurol* 2009, **512**:514-528.
- Craxton M: Genomic analysis of synaptotagmin genes. *Genomics* 2001, **77**:43-49.
- Craxton M: Synaptotagmin gene content of the sequenced genomes. *BMC Genomics* 2004, **5**:43.
- HUGO Gene Nomenclature Committee. <http://www.genenames.org>.
- Mouse Genome Informatics. <http://www.informatics.jax.org>.

33. Craxton M: **Evolutionary genomics of plant genes encoding N-terminal-TM-C2 domain proteins and the similar FAM62 genes and synaptotagmin genes of metazoans.** *BMC Genomics* 2007, **8**:259.
34. Schapire AL, Voigt B, Jasik J, Rosado A, Lopez-Cobollo R, Menzel D, Salinas J, Mancuso S, Valpuesta V, Baluska F, Botella MA: **Arabidopsis synaptotagmin 1 is required for the maintenance of plasma membrane integrity and cell viability.** *Plant Cell* 2008, **20**:3374-3388.
35. Yamazaki T, Kawamura Y, Minami A, Uemura M: **Calcium-dependent freezing tolerance in Arabidopsis involves membrane resealing via synaptotagmin SYT1.** *Plant Cell* 2008, **20**:3389-3404.
36. Südhof TC: **Synaptotagmins: why so many?.** *J Biol Chem* 2002, **277**:7629-7632.
37. Schapire AL, Valpuesta V, Botella MA: **Plasma membrane repair in plants.** *Trends Plant Sci* 2009, **14**:645-652.
38. Fitch WM: **Homology a personal view on some of the problems.** *Trends Genet* 2000, **16**:227-231.
39. Koonin EV: **Orthologs, Paralogs, and Evolutionary Genomics.** *Annu Rev Genet* 2005, **39**:309-338.
40. Liao BY, Zhang J: **Null mutations in human and mouse orthologs frequently result in different phenotypes.** *Proc Natl Acad Sci USA* 2008, **105**:6987-6992.
41. Studer RA, Robinson-Rechavi M: **How confident can we be that orthologs are similar but paralogs differ?.** *Trends Genet* 2009, **25**:210-216.
42. Lynch M: *The origins of genome architecture* Sinauer Associates, Sunderland, MA 2007.
43. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P: *Molecular biology of the cell* Garland Science, New York, NY 2008.
44. Koonin EV: **Evolution of genome architecture.** *Int J Biochem Cell Biol* 2009, **41**:298-306.
45. Koonin EV: **Darwinian evolution in the light of genomics.** *Nucleic Acids Res* 2009, **37**:1011-1034.
46. **DOE Joint Genome Institute.** <http://www.jgi.doe.gov>.
47. King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzov R, Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JG, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D: **The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans.** *Nature* 2008, **451**:783-788.
48. Tordai H, Nagy A, Farkas K, Bányai L, Patthy L: **Modules, multidomain proteins and organismic complexity.** *FEBS J* 2005, **272**:5064-5078.
49. Koonin EV: **The biological big bang model for the major transitions in evolution.** *Biol Direct* 2007, **2**:21.
50. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno MA, Kamm K, Grimwood J, Schmutz J, Shapiro H, Grigoriev IV, Buss LW, Schierwater B, Dellaporta SL, Rokhsar DS: **The Trichoplax genome and the nature of placozoans.** *Nature* 2008, **454**:955-960.
51. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS: **Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization.** *Science* 2007, **317**:86-94.
52. **National Center for Biotechnology Information.** <http://www.ncbi.nlm.nih.gov>.
53. **UCSC Genome Bioinformatics.** <http://genome.ucsc.edu>.
54. Kent WJ: **BLAT—the BLAST-like alignment tool.** *Genome Res* 2002, **12**:656-664.
55. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
56. Staden R: **The Staden sequence analysis package.** *Mol Biotechnol* 1996, **6**:109-117.
57. Corpet F: **Multiple sequence alignment with hierarchical clustering.** *Nucleic Acids Res* 1988, **16**:10881-10890.
58. **EMBL-EBI.** <http://www.ebi.ac.uk/Tools/clustalw2/>.
59. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**:2947-2948.
60. **FlyBase Genetic nomenclature for *Drosophila melanogaster*.** http://flybase.org/static_pages/docs/nomenclature/nomenclature3.html.
61. **WormBase UserGuide:Nomenclature.** <http://www.wormbase.org/wiki/index.php/UserGuide:Nomenclature>.
62. **Zebrafish Nomenclature Guidelines.** http://zfin.org/zf_info/nomen.html.
63. Sprague J, Bayraktaroglu L, Clements D, Conlin T, Fashena D, Frazer K, Haendel M, Howe D, Mani P, Ramachandran S, Schaper K, Segerdell E, Song P, Sprunger B, Taylor S, Van Slyke C, Westerfield M: **The Zebrafish Information Network: the zebrafish model organism database.** *Nucleic Acids Res* 2006, **34**:D581-D585.
64. **Xenbase.** <http://www.xenbase.org/common>.
65. Bowes JB, Snyder KA, Segerdell E, Gibb R, Jarabek C, Noumen E, Pollet N, Vise P: **Xenbase: a *Xenopus* biology and genomics resource.** *Nucleic Acids Res* 2008, **36**:D761-D767.
66. **Nomenclature for naming loci, alleles, linkage groups, and chromosomes to be used in poultry genome publications and databases.** <http://projects.roslin.ac.uk/chickmap/nomenclature.html>.
67. Min SW, Chang WP, Südhof TC: **E-Syts, a family of membranous Ca2+-sensor proteins with multiple C2 domains.** *Proc Natl Acad Sci USA* 2007, **104**:3823-3828.
68. Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, She X, Bult CJ, Agarwala R, Cherry JL, DiCuccio M, Hlavina W, Kapustin Y, Meric P, Maglott D, Birtle Z, Marques AC, Graves T, Zhou S, Teague B, Potamousis K, Churas C, Place M, Herschleb J, Runnheim R, Forrest D, Amos-Landgraf J, Schwartz DC, Cheng Z, Lindblad-Toh K, Eichler EE, Ponting CP: **Lineage-specific biology revealed by a finished genome assembly of the mouse.** *PLoS Biol* 2009, **7**:e1000112.
69. **Tree families database.** <http://www.treefam.org>.
70. Ruan J, Li H, Chen Z, Coghlan A, Coin LJ, Guo Y, Hériché JK, Hu Y, Kristiansen K, Li R, Liu T, Moses A, Qin J, Vang S, Vilella AJ, Ureta-Vidal A, Bolund L, Wang J, Durbin R: **TreeFam: 2008 Update.** *Nucleic Acids Res* 2008, **36**:D735-D740.
71. Han W, Rhee JS, Maximov A, Lao Y, Mashimo T, Rosenmund C, Südhof TC: **N-glycosylation is essential for vesicular targeting of synaptotagmin 1.** *Neuron* 2004, **41**:85-99.
72. Kanno E, Fukuda M: **Increased plasma membrane localization of O-glycosylation-deficient mutant of synaptotagmin I in PC12 cells.** *J Neurosci Res* 2008, **86**:1036-1043.
73. Ferguson GD, Vician L, Herschman HR: **Synaptotagmin IV: biochemistry, genetics, behavior, and possible links to human psychiatric disease.** *Mol Neurobiol* 2001, **23**:173-1185.
74. Inoue S, Imamura A, Okazaki Y, Yokota H, Arai M, Hayashi N, Furukawa A, Itokawa M, Oishi M: **Synaptotagmin XI as a candidate gene for susceptibility to schizophrenia.** *Am J Med Genet B Neuropsychiatr Genet* 2007, **144B**:332-340.
75. Kaufmann D, Kenner O, Nurnberg P, Vogel W, Bartelt B: **In NF1, CFTR, PER3, CARS and SYT7, alternatively included exons show higher conservation of surrounding intron sequences than constitutive exons.** *Eur J Hum Genet* 2004, **12**:139-149.
76. Gustavsson N, Lao Y, Maximov A, Chuang JC, Kostromina E, Repa JJ, Li C, Radda GK, Südhof TC, Han W: **Impaired insulin secretion and glucose intolerance in synaptotagmin-7 null mutant mice.** *Proc Natl Acad Sci USA* 2008, **105**:3992-3997.
77. Gustavsson N, Wei SH, Hoang DN, Lao Y, Zhang Q, Radda GK, Rorsman P, Südhof TC, Han W: **Synaptotagmin-7 is a principal Ca2+ sensor for Ca2+-induced glucagon exocytosis in pancreas.** *J Physiol* 2009, **587**:1169-1178.
78. Fukuda M, Kanno E, Mikoshiba K: **Conserved N-terminal cysteine motif is essential for homo- and heterodimer formation of synaptotagmins III, V, VI, and X.** *J Biol Chem* 1999, **274**:31421-31427.
79. Tomes CF: **Molecular mechanisms of membrane fusion during acrosomal exocytosis.** *Spermatology* Nottingham: Nottingham University Press/Roldan ERS, Gomendio M 2007, **65**:275-292.
80. Olesen C, Nyeng P, Kalisz M, Jensen TH, Møller M, Tommerup N, Byskov AG: **Global gene expression analysis in fetal mouse ovaries with and without meiosis and comparison of selected genes with meiosis in the testis.** *Cell Tissue Res* 2007, **328**:207-221.
81. Fukuda M, Sagi-Eisenberg R: **Confusion in the nomenclature of synaptotagmins V and IX which is which?.** *Calcium Binding Proteins* 2008, **3**:1-4.

82. Maximov A, Shin OH, Liu X, Südhof TC: **Synaptotagmin-12, a synaptic vesicle phosphoprotein that modulates spontaneous neurotransmitter release.** *J Cell Biol* 2007, **176**:113-1124.
83. Jahn JE, Coleman WB: **Re-expression of tumorigenicity after attenuation of human synaptotagmin 13 in a suppressed microcell hybrid cell line.** *Int J Oncol* 2008, **32**:441-449.
84. Quintero-Rivera F, Chan A, Donovan DJ, Gusella JF, Ligon AH: **Disruption of a synaptotagmin (SYT14) associated with neurodevelopmental abnormalities.** *Am J Med Genet A* 2007, **143**:558-563.
85. Han KH, Lee UY, Jang YS, Cho YM, Jang YM, Hwang IA, Ghee JY, Lim SW, Kim WY, Yang CW, Kim J, Kwon OJ: **Differential regulation of B/K protein expression in proximal and distal tubules of rat kidneys with ischemia-reperfusion injury.** *Am J Physiol Renal Physiol* 2007, **292**:F100-F106.
86. Morris NJ, Ross SA, Neveu JM, Lane WS, Lienhard GE: **Cloning and preliminary characterization of a 121 kDa protein with multiple predicted C2 domains.** *Biochim Biophys Acta* 1999, **1431**:525-530.
87. Fukuda M, Yamamoto A: **Assay of the Rab-binding specificity of rabphilin and Noc2: target molecules for Rab27.** *Methods Enzymol* 2005, **403**:469-481.
88. Mahoney TR, Liu Q, Itoh T, Luo S, Hadwiger G, Vincent R, Wang ZW, Fukuda M, Nonet ML: **Regulation of synaptic transmission by RAB-3 and RAB-27 in *Caenorhabditis elegans*.** *Mol Biol Cell* 2006, **17**:2617-2625.
89. Kotake K, Ozaki N, Mizuta M, Sekiya S, Inagaki N, Seino S: **Noc2, a putative zinc finger protein involved in exocytosis in endocrine cells.** *J Biol Chem* 1997, **272**:29407-29410.
90. Handley MT, Burgoyne RD: **The Rab27 effector Rabphilin, unlike Granuphilin and Noc2, rapidly exchanges between secretory granules and cytosol in PC12 cells.** *Biochem Biophys Res Commun* 2008, **373**:275-281.
91. Orita S, Sasaki T, Naito A, Komuro R, Ohtsuka T, Maeda M, Suzuki H, Igarashi H, Takai Y: **Doc2: a novel brain protein having two repeated C2-like domains.** *Biochem Biophys Res Commun* 1995, **206**:439-448.
92. Fukuda M, Mikoshiba K: **Doc2gamma, a third isoform of double C2 protein, lacking calcium-dependent phospholipid binding activity.** *Biochem Biophys Res Commun* 2000, **276**:626-632.
93. Orita S, Naito A, Sakaguchi G, Maeda M, Igarashi H, Sasaki T, Takai Y: **Physical and functional interactions of Doc2 and Munc13 in Ca²⁺-dependent exocytotic machinery.** *J Biol Chem* 1997, **272**:16081-16084.
94. DiAntonio A, Parfitt KD, Schwarz TL: **Synaptic transmission persists in synaptotagmin mutants of *Drosophila*.** *Cell* 1993, **73**:1281-1290.
95. Nonet ML, Grundahl K, Meyer BJ, Rand JB: **Synaptic function is impaired but not eliminated in *C. elegans* mutants lacking synaptotagmin.** *Cell* 1993, **73**:1291-1305.
96. Geppert M, Goda Y, Hammer RE, Li C, Rosahl TW, Stevens CF, Südhof TC: **Synaptotagmin I: a major Ca²⁺ sensor for transmitter release at a central synapse.** *Cell* 1994, **79**:717-727.
97. Poopatanapong A, Teramitsu I, Byun JS, Vician LJ, Herschman HR, White SA: **Singing, but not seizure, induces synaptotagmin IV in zebra finch song circuit nuclei.** *J Neurobiol* 2006, **66**:1613-1629.
98. Wong JL, Koppel DE, Cowan AE, Wessel GM: **Membrane hemifusion is a stable intermediate of exocytosis.** *Dev Cell* 2007, **12**:653-659.
99. Katsuyama Y, Matsumoto J, Okada T, Ohtsuka Y, Chen L, Okada H, Okamura Y: **Regulation of synaptotagmin gene expression during ascidian embryogenesis.** *Dev Biol* 2002, **244**:293-304.
100. Imai JH, Meinertzhagen IA: **Neurons of the ascidian larval nervous system in *Ciona intestinalis*: I. Central nervous system.** *Comp Neurol* 2007, **501**:316-334.
101. Imai JH, Meinertzhagen IA: **Neurons of the ascidian larval nervous system in *Ciona intestinalis*: II. Peripheral nervous system.** *Comp Neurol* 2007, **502**:335-352.
102. Matsumoto J, Katsuyama Y, Ohtsuka Y, Chen L, Lemaire P, Okamura Y: **Functional analysis of synaptotagmin gene regulatory regions in two distantly related ascidian species.** *Dev Growth Differ* 2008, **50**:543-552.
103. Mathews EA, Mullen GP, Crowell JA, Duerr JS, McManus JR, Duke A, Gaskin J, Rand JB: **Differential expression and function of synaptotagmin 1 isoforms in *Caenorhabditis elegans*.** *Mol Cell Neurosci* 2007, **34**:642-652.
104. Reenan RA: **Molecular determinants and guided evolution of species-specific RNA editing.** *Nature* 2005, **434**:409-413.
105. Lv J, Yang Y, Yin H, Chu F, Wang H, Zhang W, Zhang Y, Jin Y: **Molecular determinants and evolutionary dynamics of wobble splicing.** *Mol Biol Evol* 2009, **26**:1081-1092.
106. Dinger ME, Pang KC, Mercer TR, Mattick JS: **Differentiating protein-coding and noncoding RNA: challenges and ambiguities.** *PLoS Comput Biol* 2008, **4**:e1000176.
107. Fukuda M, Mikoshiba K: **A novel alternatively spliced variant of synaptotagmin VI lacking a transmembrane domain. Implications for distinct functions of the two isoforms.** *J Biol Chem* 1999, **274**:31428-31434.
108. Nakhost A, Houeland G, Blandford VE, Castellucci VF, Sossin WS: **Differential regulation of transmitter release by alternatively spliced forms of synaptotagmin I.** *J Neurosci* 2003, **23**:6238-6244.
109. Bagalá C, Kolev V, Mandinova A, Soldi R, Mouta C, Graziani I, Prudovsky I, Maciag T: **The alternative translation of synaptotagmin 1 mediates the non-classical release of FGF1.** *Biochem Biophys Res Commun* 2003, **310**:1041-1047.
110. Nakhost A, Houeland G, Blandford VE, Castellucci VF, Sossin WS: **Identification and characterization of a novel C2B splice variant of synaptotagmin I.** *J Neurochem* 2004, **89**:354-363.
111. Monterrat C, Boal F, Grise F, Hémar A, Lang J: **Synaptotagmin 8 is expressed both as a calcium-insensitive soluble and membrane protein in neurons, neuroendocrine and endocrine cells.** *Biochim Biophys Acta* 2006, **1763**:73-81.
112. Gauthier BR, Duhamel DL, Iezzi M, Theander S, Saltel F, Fukuda M, Wehrle-Haller B, Wollheim CB: **Synaptotagmin VII splice variants alpha, beta, and delta are expressed in pancreatic beta-cells and regulate insulin exocytosis.** *FASEB J* 2008, **22**:194-206.
113. de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Fröhlich F, Walther TC, Mann M: **Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast.** *Nature* 2008, **455**:1251-1254.
114. Burré J, Volkmandt W: **The synaptic vesicle proteome.** *J Neurochem* 2007, **101**:1448-1462.
115. Cox HD, Thompson CM: **Purification and proteomic analysis of synaptic vesicles.** *Methods Mol Biol* 2008, **432**:259-274.
116. Signorovitch AY, Dellaporta SL, Buss LW: **Molecular signatures for sex in the Placozoa.** *Proc Natl Acad Sci USA* 2005, **102**:15518-15522.
117. Pearce VB, Voigt O: **Field biology of placozoans (*Trichoplax*): distribution, diversity, biotic interactions.** *Integrative and Comparative Biology* 2007, **47**:677-692.
118. Marlow HQ, Srivastava M, Matus DQ, Rokhsar D, Martindale MQ: **Anatomy and development of the nervous system of *Nematostella vectensis*, an anthozoan cnidarian.** *Develop Neurobiol* 2009, **69**:235-254.
119. Denes AS, Jékely G, Steinmetz PRH, Raible F, Snyman H, Prud'homme B, Ferrier DEK, Balavoine G, Arendt D: **Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria.** *Cell* 2007, **129**:277-288.
120. Hill DS, Savage RM: **Evolution, development and ecology of *Capitella* sp. I: a waxing model for polychaete studies.** *Annelids in modern biology* Hoboken, New Jersey: John Wiley & Sons; 2009, 88-115.
121. Meyer NP, Seaver EC: **Neurogenesis in an annelid: characterisation of brain neural precursors in the polychaete *Capitella* sp. I.** *Dev Biol* 2009, **335**:237-252.
122. Howe D, Costanzo M, Fey P, Gojbori T, Hannick L, Hide W, Hill DP, Kania R, Schaeffer M, St Pierre S, Twigger S, White O, Yon Rhee S: **Big data: The future of biocuration.** *Nature* 2008, **455**:47-50.

doi:10.1186/1471-2164-11-37

Cite this article as: Craxton: A manual collection of *Syt*, *Esy*, *Rph3a*, *Rph3al*, *Doc2*, and *Dblc2* genes from 46 metazoan genomes - an open access resource for neuroscience and evolutionary biology. *BMC Genomics* 2010 **11**:37.