ARTICLE
Physiologists turned Geneticists: Identifying transcripts and genes for neuronal function in the Marbled Crayfish, *Procambarus virginalis*

Wolfgang Stein, Saisupritha Talasu, Andrés Vidal-Gadea & Margaret L. DeMaegd
School of Biological Science, Illinois State University, Normal, IL 61790.

The number of undergraduate researchers interested in pursuing neurophysiological research exceeds the research laboratory positions and hands-on course experiences available because these types of experiments often require extensive experience or expensive equipment. In contrast, genetic and molecular tools can more easily incorporate undergraduates with less time or training. With the explosion of newly sequenced genomes and transcriptomes, there is a large pool of untapped molecular and genetic information which would greatly inform neurophysiological processes. Classically trained neurophysiologists often struggle to make use of newly available genetic information for themselves and their trainees, despite the clear advantage of combining genetic and physiological techniques. This is particularly prevalent among researchers working with organisms that historically had no or only few genetic tools available. Combining these two fields will expose undergraduates to a greater variety of research approaches, concepts, and hands-on experiences. The goal of this manuscript is to provide an easily understandable and reproducible workflow that can be applied in both lab and classroom settings to identify genes involved in neuronal function. We outline clear learning objectives that can be acquired by following our workflow and assessed by peer-evaluation. Using our workflow, we identify and validate the sequence of two new Gamma Aminobutyric Acid A (GABAA) receptor subunit homologs in the recently published genome and transcriptome of the marbled crayfish, *Procambarus virginalis*. Altogether, this allows undergraduate students to apply their knowledge of the processes of gene expression to functional neuronal outcomes. It also provides them with opportunities to contribute significantly to physiological research, thereby exposing them to interdisciplinary approaches.

Key words: Undergraduate, peer-mentoring, GABA, Marmorkrebs, neurophysiology, gene annotation, decapod, crustacean

The thoroughly studied nervous system of decapod crustaceans such as lobsters, crabs and crayfish has been a workhorse for a large segment of the neurobiology community for many decades (Derby and Theil, 2014; Johnson et al., 2014), in the same way *Drosophila* and *C. elegans* are to geneticists and developmental biologists. Electrophysiological experiments in the crustacean nervous system have yielded many fundamental discoveries such as electrical and inhibitory synapses (Farca Luna et al., 2010; Jirikowski et al., 2010), Na/K pump (Skou, 1957), visual processing and lateral inhibition (Zieger et al., 2013; van Oosterhout et al., 2014), presynaptic inhibition (Soedarini et al., 2012), neuromodulator actions (Stein, 2009; Nusbaum and Blitz, 2012), coordination of neural circuits (Mulloney and Smarandache-Wellmann, 2012), and network dynamics (Nadim and Bucher, 2014). Studies of the morphological and functional properties of the nervous system of decapod crustaceans also revealed several neurobiological principles, including the role of GABA as inhibitory transmitter (Bowery and Smart, 2006), the generation of rhythmic motor activity (Marder, 2000; Nusbaum and Beenhakker, 2002; Selverston et al., 2009; Stein, 2017), and the control and selection of stereotyped behaviors by modulatory command neurons (Edwards et al., 1999; Nusbaum and Beenhakker, 2002; Marder and Bucher, 2007; Stein, 2009; Harris-Warrick, 2011). One of the reasons for this success story is the relatively easy access to large neurons with known connectivity. By now, due to the availability of inexpensive laboratory equipment and simplified protocols (Land et al., 2001; Johnson et al., 2007; Johnson et al., 2014; Nesbit et al., 2015; Weller et al., 2015), behavioral and electrophysiological studies on decapod crustaceans can be carried out by undergraduate researchers and many of the initial groundbreaking experiments have made it into lab courses for undergraduates and high schools (Wyttenbach et al., 2018). This has been particularly true for experiments on various species of crayfish because of their easy husbandry, availability from commercial vendors, and potential for cross-species comparisons to identify general principles of motor control (Harris-Warrick, 2011; Marzullo and Gage, 2012).

However, to fully understand how nervous systems generate behaviors, genetic and molecular insights into their workings are paramount. Many classical systems, including decapod crustaceans, lag behind in established molecular and genetic approaches to study the subcellular underpinnings of neuronal function and circuit dynamics. This is, in part, because classically trained electrophysiologists lack the time or confidence to develop the necessary protocols (see Dearborn et al., 1998; Yazawa et al., 2005; Posiri et al., 2013). This extends into classroom settings as well, such that physiology lab courses remain almost entirely separate from molecular and genetic ones.

Our goal is to establish an easily understandable and reproducible tutorial undergraduates can follow to apply genetic methods in a neurophysiological framework. This tutorial is intended to be broad enough to be incorporated into research labs and classroom environments. Our main educational objective is that undergraduates explore the process of gene expression and consider its consequences for neuronal activity and physiology.

Here, we make use of the recently published genomes and transcriptomes (Gutekunst et al., 2018) of the Marbled crayfish (*Procambarus virginalis,* Figure 1A) to outline protocols undergraduate and high school students can utilize to identify genes of interest for neurophysiological studies. Marbled crayfish are all-female triploids (Vogt et al., 2015) that produce genetically uniform offspring via apomictic parthenogenesis (Martin et al., 2007) - the development of oocytes without fertilization and meiosis (Seitz et al., 2005; Martin et al., 2007; Vogt, 2010, 2011). Its relatively short generation time and easy husbandry (Vogt et al., 2004) make this species ideal for inexpensive and streamlined research approaches and accordingly marbled crayfish have been used to study morpho-functional relationships (Vogt et al., 2004; Polanska et al., 2007; Vogt et al., 2008), neurobiology (Vilpoux et al., 2006; Fabritius-Vilpoux et al., 2008; Rieger and Harzsch, 2008), cell and body development (Seitz et al., 2005; Alwes and Scholtz, 2006; Jirikowski et al., 2010), epigenetics (Schiewek et al., 2007; Vogt et al., 2008, 2009), stem cell biology (Vogt, 2010), behavior (Vogt et al., 2008; Farca Luna et al., 2009), biogerontology (Vogt, 2010), biochronology (Farca Luna et al., 2010), as well as toxicology (Vogt, 2007; Rubach et al., 2011), ecology (Jones et al., 2009; Chucholl and Pfeiffer, 2010), and evolutionary biology (Sintoni et al., 2007).

We established this tutorial based on a course where students earn credit for participating in a solely research-based education. In this course, undergraduate students carry out independent research projects. Weekly, they meet with faculty or graduate students for research instruction and background knowledge, proposed hypotheses and mentorship. They are assessed by presenting predictions, and research results and conclusions. While this course did not include dedicated lecture periods, our tutorial can be adapted to accompany several weeks of in-depth lectures. To guide the overall structure of such courses we outline generally how to identify genes and transcripts homologous to previously published genes and how to ensure the validity of the identified sequences by assessing their presumed function. Then, we create a detailed workflow for efficient gene identification in the marbled crayfish that includes quality control processes and assessment of the learning objectives. We explain and discuss the approaches and workflow using the example of two Gamma Aminobutyric Acid A (GABAA) receptor subunits.

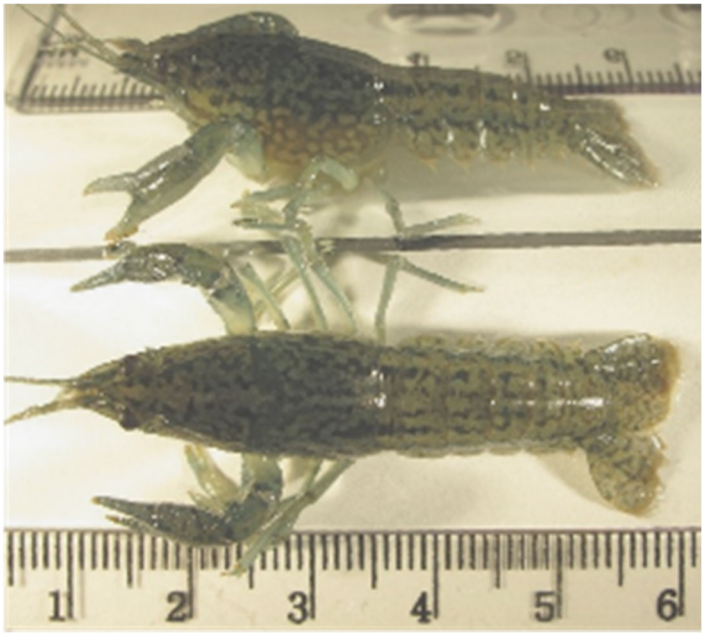

Figure 1. Dorsal and lateral view of a marbled crayfish.

MATERIALS

Three web-based search engines with free access were used: The Basic Local Alignment Search Tool of the National Center for Biotechnology (NCBI): https://www.ncbi.nlm.nih.gov/genbank and its conserved domain database: https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?, the Universal Protein Resource (UniProt) of the UniProt consortium: http://www.uniprot.org, and the Genome Portal for Marbled Crayfish at the German Cancer Research Center: http://marmorkrebs.dkfz.de.

Figures were prepared with Coreldraw (version X7, Corel Corporation, Ottawa, ON, Canada). Search results were collected with Google Sheets and Google Docs, although any spreadsheet and word processor can be used.

LEARNING OBJECTIVES

The three main learning objectives for students using this tutorial are to: (1) Develop a testable hypothesis about a physiological phenomenon and understand how an interdisciplinary approach of physiology and genetics can be stronger than either field alone. (2) Demonstrate their understanding of the process of gene expression, specifically transcription from DNA to RNA. (3) Acquire practical skills necessary for gene identification and validation.

The tutorial we outline was specifically designed to allow students to conceptualize the interconnectedness of genetics, molecular, and physiological neuroscience. It follows recently published suggestions outlined for student-driven genome annotation (Hosmani et al., 2019). Our tutorial uses Marbled Crayfish for an example species, but core concepts and approaches can be applied to any genetically tractable organism. Best learning outcomes in larger classroom settings will be achieved by students with prior experience in genetics and physiology. These will typically be junior or senior level undergraduates who have taken basic courses in genetics, or human or animal physiology. In smaller settings, such as research courses with individualized instruction for each student, even less advanced students can achieve the learning goals. This is best illustrated by one of the authors of this tutorial, who took this research class as a high school senior (S.T.) and has

successfully demonstrated a deep understanding of hypothesis-driven gene curation (Talasu and Stein, 2019).

CLASSROOM IMPLEMENTATION

Organizing an Applied Genetics Course Focused on Neurophysiology

Figure 2 outlines the general progression of gene identification within a physiological framework and highlights the key components (light colored boxes) of gene identification and the specific methods associated with them. The outline is broad enough to be applied in a variety of classroom settings, small and large. Here, each lesson should focus on a key component (light colored boxes in Figure 2) with additional time allocated to discussing the surrounding principles and technical approaches. Before moving on to the next components, learning outcomes can be assessed by having students apply the learned principles to their individualized projects. Ideally, students will build peer-mentoring networks by discussing the reasoning behind their approaches and their research outcomes.

For example, lectures pertaining to "Identifying neurophysiological process that could benefit from understanding the underlying genes and proteins" (Figure 2, top) could focus on how the diversity of genes expressed, the location of gene expression, and timing of gene expression can affect neuron activity and physiology. The culmination of this key component could end with students developing their own hypotheses and specific predictions about a physiological phenomenon which in turn will guide their gene identification and ultimately be the topic of their final assessment. To promote peer-mentoring networks, students should discuss their hypotheses and specific predictions with a partner and make revisions before submitting them for assessment. Using this course structure, the class can be taught entirely remotely. Remote lectures can be given to the entire class before moving into break-out rooms for peer discussion. This makes this lab course viable for semesters where students are not on campus.

Organizing a Collaborative Effort to Identify Marbled Crayfish Genes

While our tutorial can be adapted to classroom settings, we had great success with individualized research education and providing undergraduate students with hands-on experiences in a neurophysiology research lab. The limiting factor here is often the time and resource commitment and lack of guidance on how to train undergraduates. Our tutorial can be pursued by a single student researcher. However, we suggest creating a peer-mentoring network (Hosmani et al., 2019) that connects new and experienced annotators. This provides new students with experienced leaders who can guide their scientific approach, discuss obtained results, and troubleshoot more challenging processes. Additionally, experienced students are able to test and solidify their knowledge about gene expression and practice disseminating that knowledge to others. By discussing their findings and through collaborative assessment, students will reconsolidate their understanding of how genes and their functions are identified, and how genes contribute to neurons and their physiology. Additionally, working in teams also ensures that the annotations created are accurate and aligned with the goals of the study in accordance with the learning objectives. This will also provide students with an opportunity to contribute significantly to the larger study and will motivate them to be more actively involved in the project.

To facilitate the peer-mentoring network, we suggest grouping student researchers into teams that focus on a single pathway of the gene annotation protocol. We created a workflow (Figure 3) that helps navigate the major questions that guide each step of the protocol. For example, students could be split into two teams, one starting by blasting the gene of interest against the transcriptome (Team Transcriptome) and the other by blasting the gene of

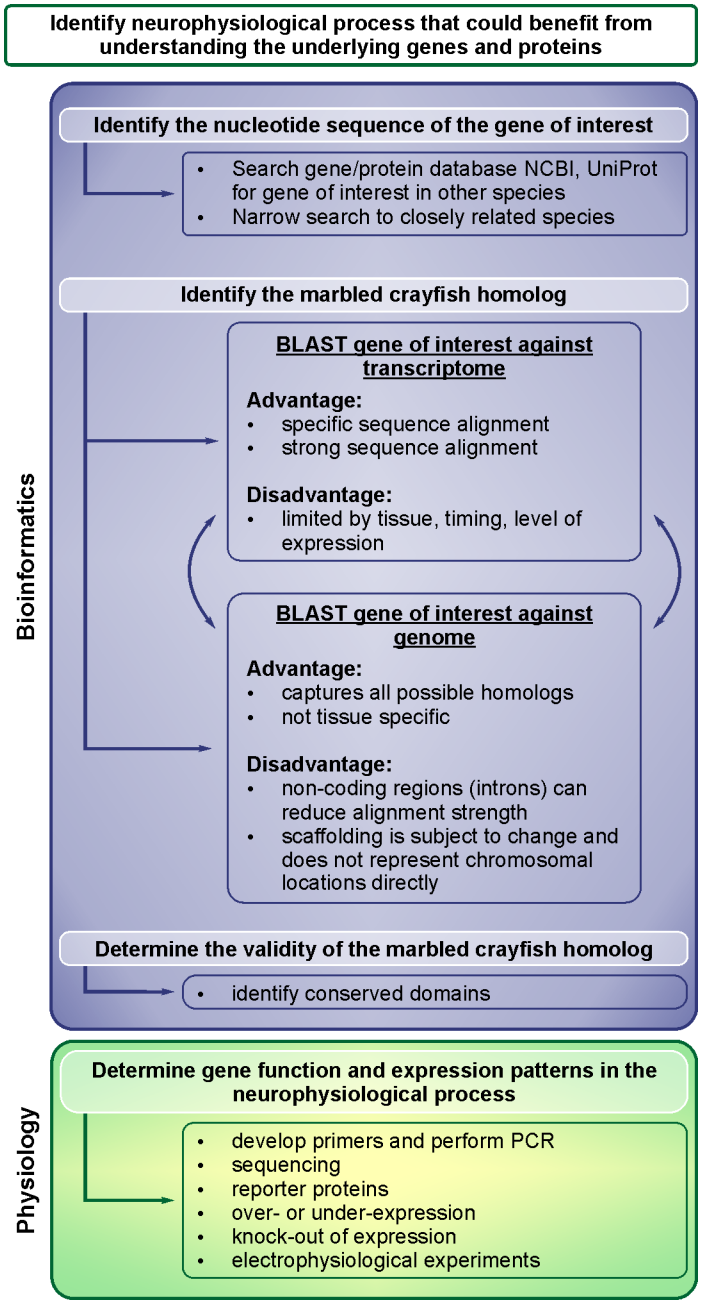

Figure 2. Flowchart to direct the combination of bioinformatics and physiological experiments.

interest against the genome (Team Genome). Both teams should follow the workflow until they have identified both transcripts and genome scaffolds and the conserved domains. At this point, the two teams will assess their results by comparing between teams to determine whether both have arrived at the same result, despite using different pathways of the flow chart. This provides iterative evaluation by peers and an opportunity to refine the annotated gene models before beginning a collaborative presentation that can be expert reviewed by senior scientists. This is a useful learning experience for the students and allows them to create a list of expected questions they may receive from the senior scientists and to generate questions they need to ask.

In addition to the student teams, senior researchers should regularly assess student understanding of the process and analyses they are currently performing by asking questions that probe the motivations and expectations behind the particular experiment. Questions could be directed toward the broad scientific merit such as, "How does the gene of interest contribute to neuronal function?" or be more specific to the task at hand such as, "Based on the gene of interest, what do you expect the conserved domain to be?"

Our outlined protocol is completely free and open-sourced, making it possible to be implemented not only in a heavily resourced lab setting, but also in laboratory classes or even remotely in the students' homes. Meetings among students and with senior scientists can be held through video conferences, and thus accommodate circumstances that prevent students from physically participating.

Developing a Hypothesis and Choosing Genes of Interest

Identifying and annotating genes can be a long and daunting task, meaning the overarching physiological goal can get lost in the details of the protocols Thus, it is important to begin by developing questions about a physiological phenomenon and ensuring that the resulting hypotheses and predictions address these questions. Our approach was to develop a workflow that could be referred to throughout the gene identification process (Figure 3). The workflow directs the thought process of the researcher to keep the biological question in mind.

The first step in a manual gene annotation is the selection of the gene of interest. This obviously depends on the research interests of the laboratory or the student. An assessment of background literature, or previously established physiological experiments should guide the research interest. Genes should then be prioritized according to the aims of the project and specific hypothesis being tested. Depending on the research question, a good strategy for prioritizing genes is to compile an initial list of gene families from which students can choose or which can be ranked by the team leader (e.g., a graduate student). This will help team members to understand the significance of their annotations and of the respective genes in neurobiological processes. For the purpose of this manuscript, we chose the GABAA receptor subunits. GABA's role in synaptic transmission and neuronal inhibition

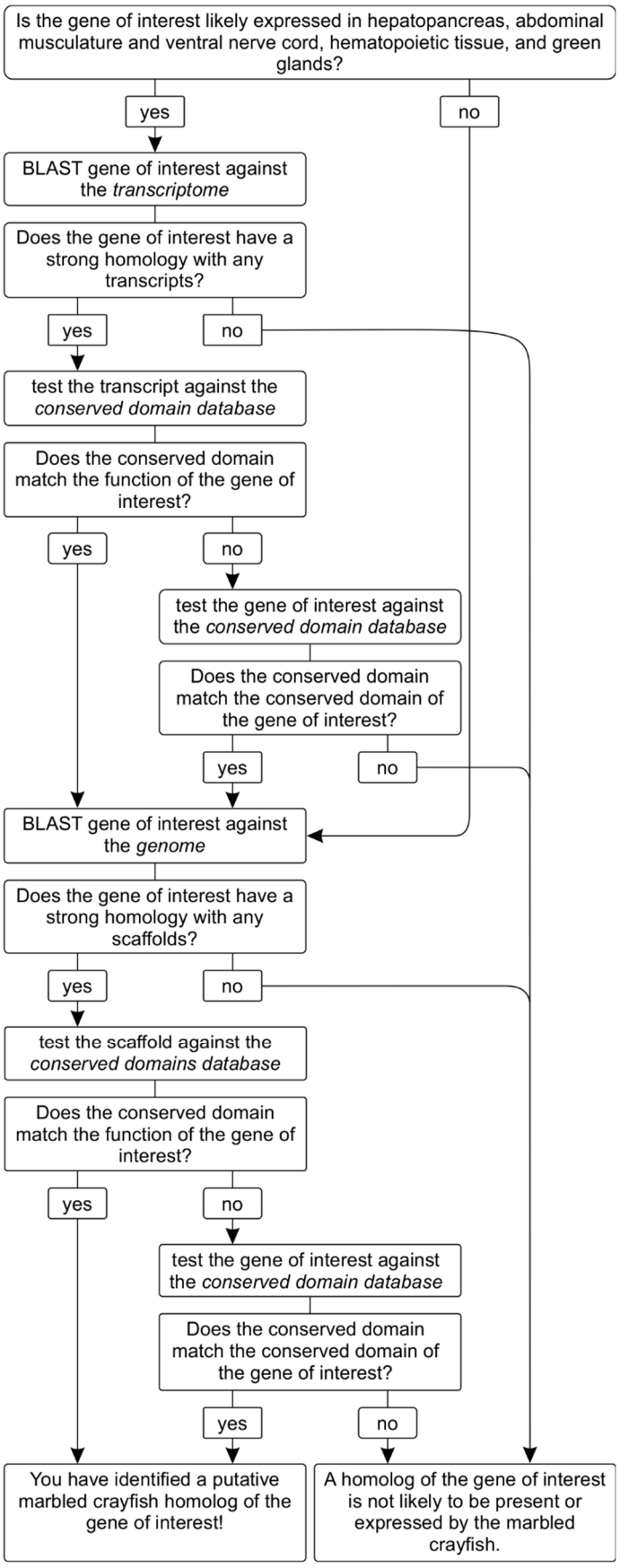

Figure 3. Workflow to identify marbled crayfish homologs of genes of interest.

were first identified in crustaceans (Kuffler and Edwards, 1958), but have since been identified in a wide variety of other invertebrates and vertebrate species (Florey, 1991). Generally, these receptors are split into two classes which correspond to fast acting ionotropic receptors and slower metabotropic receptors (Jembrek and Vlainic, 2015). In adult vertebrate nervous systems these generally have inhibitory functions, but this generalized characterization is less reliable in invertebrate species and has also been challenged in vertebrates (Stein and Nicoll, 2003). Our goal here is to identify the marbled crayfish homolog of GABAA receptor subunits, their transcripts, and potential isoforms. However, the same process can be used with other genes of interest, by replacing our query results with other genes of interest.

Both the marbled crayfish genome (Gutekunst et al., 2018) and transcriptome are freely available. Because the transcriptome contains only the expressed genes and typically with fewer gaps (higher coverage), it is a good starting point. We will thus begin with searching for GABAA receptor homologs in the transcriptome. However, if the gene of interest is unlikely to be expressed in the tissues used for generating the transcriptome (hepatopancreas, abdominal musculature and ventral nerve cord, hematopoietic tissue, and green glands) or is not expected to be expressed in the developmental stages used, then starting with the genome and following up with assessment of the transcriptome is the better option.

Identify Genes of Interest From Previously Published Sequences

The first step in any gene curation is to determine previously published gene sequences for the gene of interest. For this, we made use of the NCBI database. However, there are alternative databases such as UniProt, that provide access to similar and sometimes additional collections of sequenced genes and proteins. These databases contain large numbers of fully or partially sequenced genes, transcripts, and proteins from numerous species. The first step is to search for a keyword that describes the gene of interest. We searched for "GABA receptor".

The NCBI database returns either genomic or transcript sequences of the gene of interest and allows the download of these sequences in the FASTA format. This universally accepted format can be used in all further steps of the gene identification.

The results can be narrowed by displaying only results from related phyla, taxa, or species. In our example, we narrowed the search to the taxon crustacea, which resulted in 35 total hits (Appendix Table 1) including, both partial and completely sequenced coding regions corresponding to several receptor subtypes, subunits, and associated proteins. At this point, the coordinator or student will need to decide which sequences will be of interest. We suggest that instructors ask students to answer and discuss the following questions. 1. Will identifying the gene support the electrophysiological study? A great magnitude of genes has been identified, but not all matter for every hypothesis generated. To guide students toward connecting the two fields, instructors can facilitate responses that indicate a clear connection between physiological study and genes being identified. 2. What is the quality of the sequence? To assess the quality of the sequence we considered the evolutionary distance between species, the completeness of the sequence, and the source that identified the sequence. In a classroom setting, these topics could be expanded in lecture discussions with the following concepts emerging from the discussion. Firstly, the more closely related two species are, the more homologous their coding and non-coding sequences will be, making the sequences more likely to align. Secondly most genes are hundreds to thousands of base pairs long which can make sequencing their entirety challenging. Thus, the published sequence may not represent the entire gene. Many of the genes titles in the NCBI database identify whether the sequence corresponds to the complete or partial coding sequences (cds) which can direct the user toward the longer sequence if the same gene has been sequenced multiple times. Finally, the source of the sequence should also be considered. The NCBI database also includes links to the original literature in which the sequence was published.

We focused on two ionotropic GABAA receptor subunits. For this, we selected "*Cancer borealis* GABA receptor LCCH3-like protein mRNA" (Accession No. KU986871, Northcutt et al., 2016) and "*Procambarus clarkii* GABAA receptor subunit mRNA" (Accession No. KM115031, Jimenez-Vazquez et al., 2016, Table 1). These two transcripts have previously been identified to encode distinct GABAA receptor subunits with distinct expression profiles and properties (Northcutt et al., 2016). They also represent good candidates to find homologs in marbled crayfish because they were identified in two species of decapod crustaceans and the sequences are comprised of a large number of base pairs (greater than 1000), indicative of complete coding sequences.

Sequences can be selected by clicking on the FASTA link of each individual gene, or via bulk download of FASTA nucleotide sequences by selecting several genes and clicking "Send to:" and then checking "Coding sequences," and "FASTA Nucleotide" in the drop-down window before selecting "Create file." In either case it is useful to keep a copy of all sequences of all genes of interest, in a text file, a word processor file, or a spreadsheet, because it can be both frustrating and time consuming to search for sequences repetitively.

We used a spreadsheet to record all our searches

Name	Accession	Length (basepairs)	State of sequence
Cancer borealis GABA receptor LCCH3-like protein mRNA	KU986878.1	1446	complete cds
Procambarus clarkii GABAA receptor subunit mRNA	KM115031.1	1906	complete cds

Table 1. Abbreviated results of the Nucleotide NCBI search for genes of interest. State of the sequence is given in terms of the coding sequences (cds). Accession numbers can be used to find detailed results of each transcript and associated FASTA files.

(Appendix Table 2). Additionally, the spreadsheet may be used to assess students' participation throughout the semester and give them opportunities to receive feedback to refine their searches, results, and understanding. It will also guide students through the learning objectives, and help the coordinator determine whether students have achieved the learning goals. Finally, it will serve as a base for the final project assessment.

Identification of GABAA Receptors in the Marbled Crayfish Transcriptome

To identify whether marbled crayfish express homologs to *Cancer borealis* GABA receptor LCCH3-like protein and *Procambarus clarkii* GABAA receptor subunits, we went to the Marbled Crayfish Blast Server (http://marmorkrebs.dkfz.de/wwwblast/blast/mcblast.html). There, we pasted, one at a time, the sequences from our search query above. By selecting "PVir-transcriptome (CDS)" as the database, the result of this BLAST search (Altschul et al., 1997) will be a list of all sequences that have potential homology in the transcriptome (Table 2, see Appendix Table 3 for all results). We found 49 sequences producing significant alignments (hits) to the *Cancer borealis* GABA receptor LCCH3-like protein, and 12 hits to the *Procambarus clarkii* GABAA receptor subunit. The quality of the alignment is described with two mathematical descriptions. The first is the "Score" in bits, which describes how similar the query sequences are to the marbled crayfish transcript sequences. The score considers the number of identical base pairs, substituted base pairs, and gaps in base pairs in its assessment. This is then normalized to the statistical properties of the scoring system, which allows for scores of different alignment searches to be compared with one another. Larger scores indicate more identical sequences and better alignments. The second assessment is the "Expectation value," or "E value." The E value represents the possibility of finding different alignments with equivalent or better scores by random chance. Lower E values represent more significant scores and better alignments. While there is no set cut-off point to determine whether the sequences are sufficiently aligned, query sequences that can be found more than once without being related to the coding sequence in question will result in E values larger than one. E values that differ much from zero may thus indicate the presence of "false positives" and that the sequence is likely to be aligned at multiple unrelated locations.

It is important to note that these assessments depend on the length of the query sequence. Meaning that the maximum value of the score, and the validity of the E value depends on the number of base pairs in your query sequence. Longer query sequences have the potential to reach higher scores and are less likely to align to other sequences by random chance, increasing confidence in the obtained results.

In our case, a single transcript hit in both searches had much higher scores, and lower E values than other transcript hits, indicating that these marbled crayfish transcripts have strong and specific alignments to the genes of interest. Specifically, transcript TR 18728 was the top hit for the *Cancer borealis* GABA receptor LCCH3-like protein, while TR 2556 was the top hit for the *Procambarus clarkii* GABAA receptor subunit. The identification of two different transcripts could be a result of either a) two different genes encoding these transcripts and thus two distinct GABAA receptor subunits or b) there is splice variants of a single GABAA receptor gene. To determine which of these options is more likely, the location and sequence alignment of these transcripts within the genome needs to be assessed. We describe this process further down in the section titled "Identification of the genomic sequence of the marbled crayfish GABAA receptor subunits." However, before describing this step, we wanted to confirm whether our identified transcripts not just showed sequence but also functional homology to GABAA receptors. For this, we recorded the transcript number, score, and E value, and the transcript sequences. To obtain the transcript sequence, the full transcriptome can be downloaded from (http://marmorkrebs.dkfz.de/downloads/, Gutekunst et al., 2018). Once downloaded, the transcript can be opened in a text file reader, and the transcript number can be searched for and the sequence can be pasted into the Google sheet (Note that if the transcript starts with zero, as in 02556, the initial zero should be discarded).

In contrast to what we observed, there may be several transcripts that have similar scores and E values. In this case, recording several homologous transcripts can be appropriate. Similar scores and E values can result when only a portion of the query sequence aligns with the marbled crayfish transcriptome. Partial alignments between the query sequence and transcriptome can happen for a variety of reasons and a lecture period or in-person discussion should be devoted to considering them. One option would

Name	Transcript	Score (bits)	E value
Cancer borealis GABA receptor LCCH3-like protein mRNA	18728	846	0
Cancer borealis GABA receptor LCCH3-like protein mRNA	12698	58	2E-7
Cancer borealis GABA receptor LCCH3-like protein mRNA	14362	40	0.039
Procambarus clarkii GABAA receptor subunit mRNA	2556	2839	0
Procambarus clarkii GABAA receptor subunit mRNA	18728	52	1E-5
Procambarus clarkii GABAA receptor subunit mRNA	9841	46	8E-4

Table 2. Marbled crayfish transcripts with clear sequence alignments to either the *Cancer borealis* GABA receptor LCCH3-like receptor subunit (top) or *Procambarus clarkii* GABAA receptor subunit (bottom). Best three hits only.

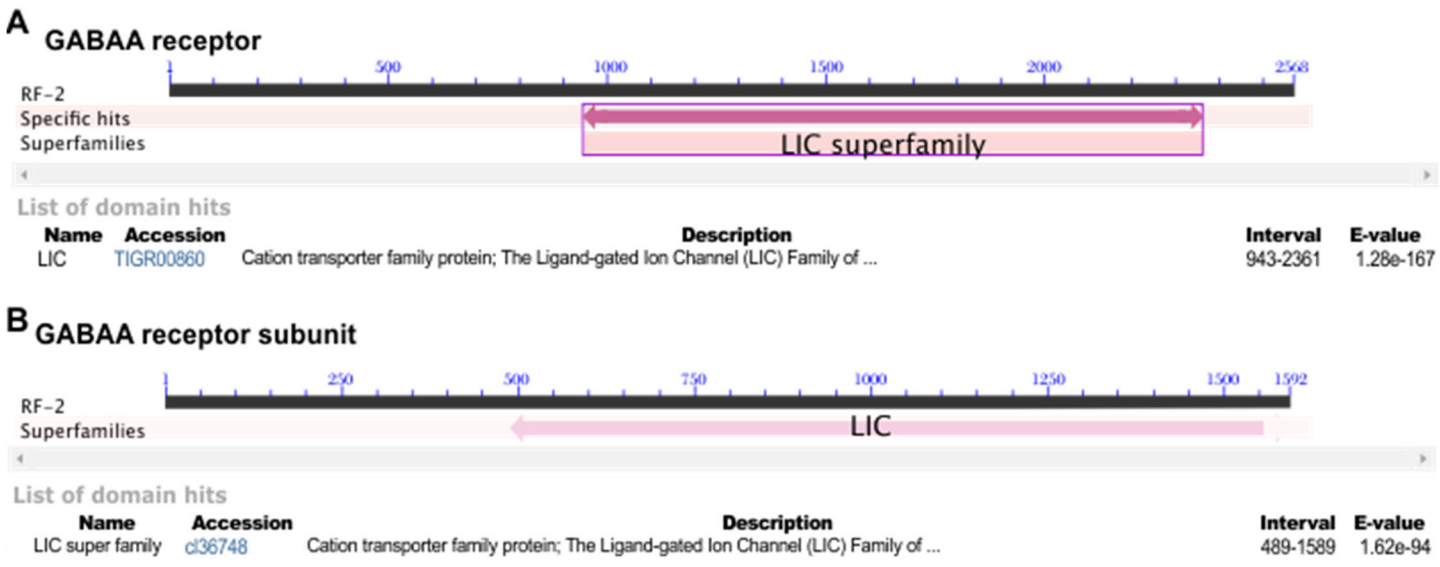

Figure 4. Conserved domains results of the two identified transcripts *(A)* TR 18728 (B) TR 2556. Image taken directly from the NCBI Conserved Domain database, with permission (Marchler-Bauer and Bryant, 2004; Marchler-Bauer et al., 2011, 2015, 2017; Lu et al., 2020).

be to discuss whether the query sequence encodes a component of the protein that is involved in a broader function within the protein family or a function specific to a subtype of that family. For example, a sequence may encode a subunit common to all GABA receptors as opposed to one specific to GABAA receptors. Additionally, if the query sequence encodes a subunit that is incorporated as one of several heteromeric subunits (present, for example, in TRP channels; Palovcak et al., 2015) Finally, if the two sequences have diverged with evolutionary distance between species then they may have only partial sequence alignment.

Using the examples described above, associated lectures could include a discussion about the major domains common to proteins of the same class, followed by a focus on functional domains and protein subtypes with different functions. Follow-up lectures should also address gene mutation and evolution, and potential functional changes arising from them. In doing so, this would allow students to address the concept that while general receptor identify is ubiquitous across phyla, species specific differentiation results from genetic changes over time and result in different pharmacological and functional profiles.

The BLAST identifies transcripts with sequence homology and provides an analysis of how strong the sequence alignment is. However, it does not provide information about whether the alignment is to a *functionally* homologous transcript. In fact, the newly identified homologous transcript could code for a protein with a related but not identical function. Serotonin and dopamine receptors are good examples for this problem: Both proteins bind to monoamines, and accordingly they have much overlap in the sequences for their binding sites, but functionally one has a higher affinity to Dopamine, while the other prefers Serotonin (Christie et al., 2020a). Additionally, similar sequences may be present that code for a region of the protein that may not be directly associated with the function of the entire protein. For example, many receptors may activate G-proteins through similar mechanisms, despite quite distinct receptor binding domains. Ultimately the real test is to find that the protein is where it needs to be and interacts with the ligand predicted. An initial step to address this issue is to test the putative function of the sequence by assessing its "conserved domains" or functional units.

To test the transcript for conserved domains, we pasted the nucleotide sequence of the transcript into the search bar of the NCBI Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; Lu et al., 2020). Figure 4 shows the results of our searches using the marbled crayfish transcripts 2556 and 18728. The top section shows the graphical representation of the sequence alignment of our query to those in the conserved domain data base. A fully labelled description of the graphical representation is described by the NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd_help.shtml#ConciseDisplay). In brief, our entire transcript query is represented linearly as the grey bar labelled RF-2. The domains with significant hits are shown in color below, with the size and location relative to the query sequence. Larger regions indicate that more of the sequences are identical. The intensity of color represents how specific the hit is, with more intense colors representing more specific hits. In our example, both transcripts share sequences with the Ligand-gated ion channel (LIC) super family (in light pink). This super family represents ionotropic neurotransmitter receptors only found in animals. Invertebrate receptors in this family respond to acetylcholine, serotonin, glycine, glutamate, or GABA and preferentially transport cation or anions depending on the channel (Lester et al., 2004; Chen et al., 2006). Additionally, transcript 18728 has sequence similarity to a specific hit of a ligand-gated ion channel

subfamily (dark pink) whose function is primarily cation transport. The text below the graphical representation lists the domain hits in order of most similar to least, and provides the accession number, a brief description which can be found in more detail by clicking the accession number, the base pair interval over which the sequences align, and the E value associated with the alignment.

Because our interest was in identifying two GABAA receptor subunits, these results are consistent with our transcripts coding for proteins that are functionally homologous to known GABAA receptors and GABAA receptor beta subunits. In addition to the super families describing the function we expect, the conserved domain has strong sequence alignment as indicated by the small E values (TR 18728: 1.28E-167 and TR 2556: 1.62E-94).

It is possible that the conserved domain search will return results that do not match the expected protein function. However, this does not immediately indicate that the transcript is not functionally homologous with the gene of interest. Because there can be similarities in functional domains or species-specific sequences, the most similar conserved domain may not have the expected function. Here, a conserved domain search of the original gene of interest nucleotide sequence should be performed. If the results of both the gene of interest and identified marbled crayfish transcript match, the two likely have homologous functions (see Christie et al., 2020b), and the researcher can continue with the workflow. In contrast, if the results of the gene of interest do not match the marbled crayfish transcript results, the marbled crayfish transcript is highly unlikely to be functionally homologous to the gene of interest and the researcher should consider either other marbled crayfish transcripts with sequence homology identified by the BLAST or pursue other genes of interest.

For additional confirmation that the conserved domains of our identified transcripts matched those of the genes of interest, we searched for conserved domains using their nucleotide sequences as search queries. We found that the *Cancer borealis* GABA receptor LCCH3-like protein had strong sequence homology (E value:3.78E-172) with the same LIC specific hit found with the putative GABAA receptor transcript 18728, and the *Procambarus clarkii* GABAA receptor subunit had strong sequence homology (E value: 5.17E-112) to the same LIC superfamily hit as the putative GABAA receptor subunit transcript 2556. Because these results are similar to the search results of our marbled crayfish transcripts, we accepted this as confirmation of the putative function of our transcripts.

Identification of the Genomic Sequence of the Marbled Crayfish GABAA Receptor Subunits

To further characterize the homolog of the gene of interest, it may be of interest to also identify its genomic sequence. This is an imperative step for gene curation, as it provides insight into the intron and exon structure of the gene, and, by comparing to the transcriptome, allows the identification of splice variants and gene duplications. Knowledge of the genomic sequence also provides foundation for a phylogenetic comparison of the gene to its homologs in other species. Genomic analysis is also required to carry out some of the advanced approaches mentioned above. For example, promotors and transcription controlling elements that are up- or downstream of the coding sequence, or within introns, can be identified. Finally, identifying the genomic sequence allows one to create intron-spanning primers to measure gene expression levels.

The first step is to BLAST the sequence of interest against the marbled crayfish genome. Like the BLAST against the transcriptome, this BLAST tests for nucleotide sequence alignment, but in this case between the gene or transcript of interest and the genome. As a result, it will show the aligned sequences, and print the scaffold on which the homolog sequence can be found in the genome. Unlike a transcript which codes for only one protein, a scaffold can encode multiple transcripts. This is because a scaffold is a reconstruction of the genome based on overlapping sequences, or contigs. The marbled crayfish genome was acquired via a shotgun method (Gutekunst et al., 2018). In brief, this method breaks the DNA of the entire genome into smaller sequences, or reads. Using an assembler algorithm, the reads are sorted into combinations of partially overlapping sequences, again called contigs. Finally, a different scaffolder algorithm is used to organize the contigs based on their overlapping sequences into scaffolds. Because the algorithms focus on matching sequences, genes encoded near to one another can be on the same scaffold, or in a worst-case scenario, a single gene could be split onto different scaffolds or found on multiple scaffolds. The process of finding overlapping sequences and assigning scaffolds and scaffold numbers is ongoing and subject to change as more gene curation occurs.

We recommend starting gene curation with the transcriptome if it is likely that the gene of interest is expressed in the tissues used to create the transcriptome (Gutekunst et al., 2018). The reason for this is that the initial query sequence includes only the coding regions, and thus will likely result in better alignment to the transcriptome than the genome. However, if it is not likely expressed or the specific protein is not known, we instead recommend starting gene curation with the genomic sequence. Because the genome is based on DNA, it can be used to identify all possible genes and possible protein isoforms irrespective of the ultimate level, timing, or location of protein expression. In our case, we used the marbled crayfish transcript sequences we had identified above (TR 18728 putative GABAA receptor LCCH3 subunit, and TR 2556 putative GABAA β subunit). These sequences should show the highest homology to the genomic sequences because they are from the same species. The marbled crayfish genomic database is freely available at http://marmorkrebs.dkfz.de/wwwblast/blast/mcblast.html. The blast server is run by Dr. Frank Lyko's group at the German Cancer Research center (DKFZ). We copied the transcript sequences of the putative GABAA receptor subunits into the BlastN (nucleotide) search. Use of the default database (Pvir04-l1k at time of writing) provided the genome scaffold numbers, as well as the score and E value of the identified genomic sequences. The BLAST of TR 18728 returned 17 hits, however, 13 of these were likely false positives because their E value is greater than 1

S10872 conserved domain hits

Name	Member id	Accession	Interval	E Value
LGIC ECD super family	cd19006	cl28912	2435-2590	1.23E-31
LIC super family	TIGR00860	cl36748	494-685	3.11E-09
LGIC ECD super family	cd19006	cl28912	2751-2898	1.60E-22
LGIC ECD super family	cd19006	cl28912	3847-3909	2.29E-04
LIC super family	TIGR00860	cl36728	1218-1943	2.30E-53

S239177 conserved domain hits

Name	Member id	Accession	Interval	E Value
LGIC ECD super family	cd18990	cl28912	1393-1533	3.18E-07
Neur chan member super family	pfam02932	cl08379	4909-4974	2.32E-03
LGIC ECD super family	cd19008	cl28912	2-91	9.90E-14
LGIC TM super family	cd19049	cl38911	2973-2837	1.52E-12
LIC super family	TIGR00860	cl36748	2016-2141	3.09E-08

Table 3. Conserved domain results for marbled crayfish genomic homologs of genes of interest.

(Appendix Table 4). Of the remaining, scaffold S10872 had a much higher score and lower E value than all other scaffolds (score: 1552, E value: 0.0, Appendix Table 4). The BLAST of TR 2556 returned only 6 scaffolds, two of which are likely false positives. Scaffold S239177 had a much higher score and lower E value than the remaining (score: 1116, E value: 0.0). With respect to our goal to determine if the GABAA receptor subunits were part of the same GABAA receptor genes, because two independent scaffolds were identified, our data supports the possibility of two distinct putative GABAA receptor subunit genes in the marbled crayfish.

We recorded the scaffold numbers, scores, and E values in our Google sheet. Additionally, the full scaffold sequence is necessary for the creation of primers, for example, to test for the expression of the GABAA receptor subunits. In order to access the full sequence of the scaffold, we recommend creating an Apollo account by contacting the Lyko group (marmorkrebs@dkfz-heidelberg.de). In Apollo, one can easily download a FASTA file of the sequence by searching for the scaffold number and selecting "download gff and FASTA file." However, the scaffold sequence can still be acquired without access to Apollo. We suggest one of two ways to access the full genome sequence. It can be downloaded (http://marmorkrebs.dkfz.de/downloads/, Gutekunst et al., 2018) and searched for the scaffold. However, this method requires a text reader that can open large text files. Alternatively, the *Procambarus virginalis* genome can be accessed via the NCBI Sequence Set Browser (Project: MRZY01, https://www.ncbi.nlm.nih.gov/Traces/wgs/MRZY01). Here, all scaffolds can be searched for, and FASTA files can be accessed by selecting the “Contigs” tab and using the search bar, "Name," and query, “SEQ###,” where the ### represents the scaffold number. The results should provide the Accession number, name, length, and multiple download files of the sequence. In our Google sheet (Appendix 2) we include two versions of the scaffold. The first is the whole scaffold sequence, to assist in further assessments of the identified genes and the second is the aligned regions of the scaffold. Because the scaffold could contain multiple different genes, the scaffold region that aligns with the gene or transcript of interest can be used separately to confirm the putative identity of the genomic sequence.

Like for the transcriptome, the BLAST only identifies sequence homology and provides an analysis of sequence alignment strength. To test whether the found genomic sequences indeed align to a *functionally* homologous gene, we again used the conserved domains database. Table 3 shows the conserved domain hits for scaffold S10872. These hits were primarily ligand-gate ion channel super families. Additionally, within the super family, our scaffold aligned to specific members of the super family. One notable example includes the ligand-gate chloride channel homolog 3 (LCCH3, cd19006), which we anticipated based on our original gene of interest query, *Cancer borealis* GABA receptor LCCH3-like protein. S239177 also showed conserved domain hits corresponding to ligand-gated ion channels. The specific members of the super family with which it aligned, however, (cd19008) were Resistant to Dieldrin (RDL) GABAA receptor subunits. This supports our conclusion that the two transcripts encode two separate putative GABAA receptor subunits. Altogether, we identified homologous marbled crayfish transcripts and genomic sequences for both the *Cancer borealis* GABA receptor LCCH3-like protein and the *Procambarus clarkii* GABAA receptor subunit.

Relating the Bioinformatics Results Back to the Physiological Phenomenon

To attend to the final component of the gene curation flowchart (Figure 2), students should integrate their newly acquired bioinformatics knowledge with the physiology of the marbled crayfish. Students should propose experiments using their identified genes that would test the hypothesis they developed at the beginning of the course. Instructors can provide lectures about the molecular, electrophysiological, or behavioral approaches that may allow testing these hypotheses, or include a class discussion of an original research article (e.g., Spigelman et al., 2002; Zhang et al., 2003; Dernovici et al., 2007). While beyond the scope of this manuscript, there are also excellent examples for student neurophysiology approaches that could be combined with our tutorial (Marzullo and Gage, 2012; Wyttenbach et al., 2018).

For our specific example of GABAA receptors, students proposed the following potential experiments: (1) GABAA receptor transcripts (mRNA) expressed in an oocyte and tested for responsiveness to GABA; (2) GABAA receptor expression increased, reduced, or knocked out in the marbled crayfish, and GABA responsiveness assessed on the behavioral level; (3) GFP tagged to the expression of

the putative GABAA receptor hinting at neurons that are known to respond to GABA release; (4) Primers against GABAA receptors generated to test whether mRNA extracted from cells contains transcribed (putative) GABAA receptors; (5) Selective reduction of GABAA receptor expression via RNAi in the ventral nerve cord to test the contribution of the putative GABAA receptors in the crayfish escape circuit, when combined with behavioral or neurophysiological recordings (Edwards et al., 1999; Dzitoyeva et al., 2003).

ASSESSMENT AND STUDENT OUTCOMES

We implemented this tutorial in a Course in Undergraduate Research Education (CURE) – like setting. This course brings undergraduate students of all levels to active research labs and allows students to work under supervision of graduate students and faculty. The overall assessment of student outcomes is achieved through bi-weekly question-and-answer sessions between students and mentors, a continuous tracking of student's improvements throughout the course, and a final assessment through either oral or poster presentations of the research data and conclusions. Final assessments are carried out by an expert panel of faculty and graduate students, who test whether students have achieved the learning objectives.

We found that students were able to develop clear and concise hypotheses that combined both bioinformatics and neurophysiology (objective 1). They were able to then pursue these hypotheses, and with guidance develop specific predictions about their results throughout the tutorial, demonstrating their proficiency in the key concepts of gene expression (objective 2) and curation (objective 3). Finally, most students surpassed our learning objectives in that they were able to apply some of these key concepts to neurophysiological questions and approaches. A good example for this is a poster presented by one of the authors of this study, who took this course as a high school senior, but presented their data at a local conference for undergraduate research (Talasu and Stein, 2019).

Some students also requested to curate additional genes of interest, beyond the ones required. We had senior undergraduate students attempt, and succeed, in replicating results reported previously by graduate students in the lab. Assessment results were thus similar to those reported in other bioinformatics and biocuration modules (Grisham et al., 2010; Mitchell et al., 2015).

In larger classroom settings, combining content-driven lectures and hands-on curation will improve student motivation. This gives students the necessary knowledge and skills to contribute both technically and intellectually to the project, and to achieve the learning objectives. Assessment here should include verbal discussions to test student abilities to complete assigned gene identification components throughout the semester. Towards the end of the semester, students should be assessed through cumulative final project summaries and oral presentations of their hypotheses (objective 1), approaches and data (objective 2), and conclusions (objective 3). Presentations can be judged by either the instructor, or a mixed panel of student peers, instructors, and graduate students, and serve as measures to assess the student's comprehension and wholistic understanding of materials.

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

Alwes F, Scholtz G (2006) Stages and other aspects of the embryology of the parthenogenetic Marmorkrebs (Decapoda, Reptantia, Astacida). Dev Genes Evol 216:169-184.

Bowery NG, Smart TG (2006) GABA and glycine as neurotransmitters: a brief history. Br J Pharmacol 147 Suppl 1:S109-119.

Chen Y, Reilly K, Chang Y (2006) Evolutionarily Conserved Allosteric Network in the Cys Loop Family of Ligand-gated Ion Channels Revealed by Statistical Covariance Analyses. J Biol Chem 281:18184-18192.

Christie AE, Hull JJ, Dickinson PS (2020a) Assessment and comparison of putative amine receptor complement/diversity in the brain and eyestalk ganglia of the lobster, Homarus americanus. Invert Neurosci 20:7.

Christie AE, Hull JJ, Dickinson PS (2020b) In silico analyses suggest the cardiac ganglion of the lobster, Homarus americanus, contains a diverse array of putative innexin/innexin-like proteins, including both known and novel members of this protein family. Invert Neurosci 20:5.

Chucholl C, Pfeiffer M (2010) First evidence for an established Marmorkrebs (Decapoda, Astacida, Cambaridae) population in Southwestern Germany, in syntopic occurrence with Orconectes limosus (Rafinesque, 1817). Aquatic invasions 5:405-412.

Dearborn RE, Jr., Szaro BG, Lnenicka GA (1998) Microinjection of mRNA encoding rat synapsin Ia alters synaptic physiology in identified motoneurons of the crayfish, Procambarus clarkii. J Neurobiol 37:224-236.

Derby C, Thiel M (2014). The Natural History of Crustaceans, Vol. 3, Crustacean Nervous Systems and their Control of Behavior. New York: Oxford UP.

Dernovici S, Starc T, Dent JA, Ribeiro P (2007) The serotonin receptor SER-1 (5HT2ce) contributes to the regulation of locomotion in Caenorhabditis elegans. Dev Neurobiol 67:189-204.

Dzitoyeva S, Dimitrijevic N, Manev H (2003) Gamma-aminobutyric acid B receptor 1 mediates behavior-impairing actions of alcohol in Drosophila: adult RNA interference and pharmacological evidence. Proc Natl Acad Sci U S A 100:5485-5490.

Edwards DH, Heitler WJ, Krasne FB (1999) Fifty years of a command neuron: the neurobiology of escape behavior in the crayfish. Trends Neurosci 22:153-161.

Fabritius-Vilpoux K, Bisch-Knaden S, Harzsch S (2008) Engrailed-like immunoreactivity in the embryonic ventral nerve cord of the Marbled Crayfish (Marmorkrebs). Invert Neurosci 8(4):177-197.

Farca Luna AJ, Heinrich R, Reischig T (2010) The circadian biology of the marbled crayfish. Front Biosci 2:1414-1431.

Farca Luna AJ, Hurtado-Zavala JI, Reischig T, Heinrich R (2009) Circadian regulation of agonistic behavior in groups of parthenogenetic marbled crayfish, Procambarus sp. J Biol Rhythms 24:64-72.

Florey E (1991) GABA: history and perspectives. Can J Physiol Pharmacol 69:1049-1056.

Grisham W, Schottler NA, Valli-Marill J, Beck L, Beatty J (2010) Teaching bioinformatics and neuroinformatics by using free web-based tools. CBE Life Sci Educ 9:98-107.

Gutekunst J, Andriantsoa R, Falckenhayn C, Hanna K, Stein W, Rasamy J, Lyko F (2018) Clonal genome evolution and rapid invasive spread of the marbled crayfish. Nat Ecol Evol 2:567-573.

Harris-Warrick RM (2011) Neuromodulation and flexibility in Central Pattern Generator networks. Curr Opin Neurobiol 21:685-692.

Hosmani PS, Shippy T, Miller S, Benoit JB, Munoz-Torres M, Flores-Gonzalez M, Mueller LA, Wiersma-Koch H, D'Elia T, Brown SJ, Saha S (2019) A quick guide for student-driven community genome annotation. PLoS Comput Biol 15:e1006682.

Jembrek MJ, Vlainic J (2015) GABA Receptors: Pharmacological Potential and Pitfalls. Curr Pharm Des 21:4943-4959.

Jimenez-Vazquez EN, Diaz-Velasquez CE, Uribe RM, Arias JM, Garcia U (2016) Molecular cloning and expression of a GABA receptor subunit from the crayfish Procambarus clarkii. J Neurosci Res 94:190-203.

Jirikowski G, Kreissl S, Richter S, Wolff C (2010) Muscle development in the marbled crayfish--insights from an emerging model organism (Crustacea, Malacostraca, Decapoda). Dev Genes Evol 220:89-105.

Johnson BR, Hauptman SA, Bonow RH (2007) Construction of a simple

suction electrode for extracellular recording and stimulation. J Undergrad Neurosci Educ 6:A21-26.
Johnson BR, Wyttenbach RA, Hoy RR. (2014). Crustacean teaching models in Neuroscience. In: The Natural History of Crustaceans, Vol. 3, Crustacean Nervous Systems and their Control of Behavior, (Derby C, Thiel M, eds) pp 535-553. New York: Oxford University Press.
Jones JP, Rasamy JR, Harvey A, Toon A, Oidtmann B, Randrianarison MH, Raminosoa N, Ravoahangimalala OR (2009) The perfect invader: a parthenogenic crayfish poses a new threat to Madagascar's freshwater biodiversity. Biological Invasions 11:1475-1482.
Kuffler SW, Edwards C (1958) Mechanism of gamma aminobutyric acid (GABA) action and its relation to synaptic inhibition. J Neurophysiol 21:589-610.
Lester HA, Dibas MI, Dahan DS, Leite JF, Dougherty DA (2004) Cys-loop receptors: new twists and turns. Trends Neurosci 27:329-336.
Land BR, Wyttenbach RA, Johnson BR (2001) Tools for physiology labs: an inexpensive high-performance amplifier and electrode for extracellular recording. J Neurosci Methods 106:47-55.
Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A (2020) CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res 48:D265-D268.
Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32:W327-331.
Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH (2015) CDD: NCBI's conserved domain database. Nucleic Acids Res 43:D222-226.
Marchler-Bauer A, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225-229.
Marchler-Bauer A et al. (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200-D203.
Marder E (2000) Motor pattern generation. Curr Opin Neurobiol 10:691-698.
Marder E, Bucher D (2007) Understanding circuit dynamics using the stomatogastric nervous system of lobsters and crabs. Annu Rev Physiol 69:291-316.
Martin P, Kohlmann K, Scholtz G (2007) The parthenogenetic Marmorkrebs (marbled crayfish) produces genetically uniform offspring. Die Naturwissenschaften 94:843-846.
Marzullo TC, Gage GJ (2012) The SpikerBox: a low cost, open-source bioamplifier for increasing public participation in neuroscience inquiry. PLoS ONE 7:e30837.
Mitchell CS, Cates A, Kim RB, Hollinger SK (2015) Undergraduate Biocuration: Developing Tomorrow's Researchers While Mining Today's Data. J Undergrad Neurosci Educ 14:A56-65.
Mulloney B, Smarandache-Wellmann C (2012) Neurobiology of the crustacean swimmeret system. Prog Neurobiol 96:242-267.
Nadim F, Bucher D (2014) Neuromodulation of neurons and synapses. Curr Opin Neurobiol 29C:48-56.
Nesbit SC, Van Hoof AG, Le CC, Dearworth JR, Jr. (2015) Extracellular Recording of Light Responses from Optic Nerve Fibers and the Caudal Photoreceptor in the Crayfish. J Undergrad Neurosci Educ 14:A29-38.
Northcutt AJ, Lett KM, Garcia VB, Diester CM, Lane BJ, Marder E, Schulz DJ (2016) Deep sequencing of transcriptomes from the nervous systems of two decapod crustaceans to characterize genes important for neural circuit function and modulation. BMC Genomics 17:868.
Nusbaum MP, Beenhakker MP (2002) A small-systems approach to motor pattern generation. Nature 417:343-350.
Nusbaum MP, Blitz DM (2012) Neuropeptide modulation of microcircuits. Curr Opin Neurobiol 22:592-601.
Polanska MA, Yasuda A, Harzsch S (2007) Immunolocalisation of crustacean-SIFamide in the median brain and eyestalk neuropils of the marbled crayfish. Cell Tissue Res 330:331-344.
Palovcak E, Delemotte L, Klein ML, Carnevale V (2015) Comparative sequence analysis suggests a conserved gating mechanism for TRP channels J Gen Physiol 146(1):37–50.
Posiri P, Ongvarrasopone C, Panyim S (2013) A simple one-step method for producing dsRNA from E. coli to inhibit shrimp virus replication. Journal of Virological Methods 188:64-69.
Rieger V, Harzsch S (2008) Embryonic development of the histaminergic system in the ventral nerve cord of the Marbled Crayfish (Marmorkrebs). Tissue Cell 40:113-126.
Rubach MN, Crum SJ, Van den Brink PJ (2011) Variability in the dynamics of mortality and immobility responses of freshwater arthropods exposed to chlorpyrifos. Arch Environ Contam Toxicol 60:708-721.
Schiewek R, Wirtz M, Thiemann M, Plitt K, Vogt G, Schmitz OJ (2007) Determination of the DNA methylation level of the marbled crayfish: an increase in sample throughput by an optimised sample preparation. J Chromatogr B Analyt Technol Biomed Life Sci 850:548-552.
Seitz R, Vilpoux K, Hopp U, Harzsch S, Maier G (2005) Ontogeny of the Marmorkrebs (marbled crayfish): a parthenogenetic crayfish with unknown origin and phylogenetic position. J Exp Zoolog A Comp Exp Biol 303:393-405.
Selverston AI, Szucs A, Huerta R, Pinto R, Reyes M (2009) Neural mechanisms underlying the generation of the lobster gastric mill motor pattern. Front Neural Circuits 3:12-12.
Sintoni S, Fabritius-Vilpoux K, Harzsch S (2007) The engrailed-expressing secondary head spots in the embryonic crayfish brain: examples for a group of homologous neurons in Crustacea and Hexapoda? Dev Genes Evol 217:791-799.
Skou JC (1957) The influence of some cations on an adenosine triphosphatase from peripheral nerves. Biochim Biophys Acta 23:394-401.
Spigelman I, Li Z, Banerjee PK, Mihalek RM, Homanics GE, Olsen RW (2002) Behavior and physiology of mice lacking the GABAA-receptor delta subunit. Epilepsia 43 Suppl 5:3-8.
Soedarini B, Klaver L, Roessink I, Widianarko B, van Straalen NM, van Gestel CA (2012) Copper kinetics and internal distribution in the marbled crayfish (Procambarus sp.). Chemosphere 87:333-338.
Stein V, Nicoll RA (2003) GABA generates excitement. Neuron 37:375-378.
Stein W (2009) Modulation of stomatogastric rhythms. J Comp Physiol A 195:989-1009.
Stein W (2017) Stomatogastric Nervous System. In: Oxford Research Encyclopedia, New York: NY: Oxford University Press. Available at http://dx.doi.org/10.1093/acrefore/9780190264086.013.153.
Talasu S, Stein W (2019) Identifying Dopamine Receptor Genes and Transcription Marbled Crayfish. In: Faculty Publications – Biological Sciences 30Normal, IL: Illinois State University. Available at https://ir.library.illinoisstate.edu/fpbiosci/30.
van Oosterhout F, Goitom E, Roessink I, Lurling M (2014) Lanthanum from a modified clay used in eutrophication control is bioavailable to the marbled crayfish (Procambarus fallax f. virginalis). PLoS ONE 9:e102410.
Vilpoux K, Sandeman R, Harzsch S (2006) Early embryonic development of the central nervous system in the Australian crayfish and the Marbled crayfish (Marmorkrebs). Dev Genes Evol 216:209-223.
Vogt G (2007) Exposure of the eggs to 17alpha-methyl testosterone reduced hatching success and growth and elicited teratogenic effects in postembryonic life stages of crayfish. Aquat Toxicol 85:291-296.
Vogt G (2010) Suitability of the clonal marbled crayfish for biogerontological research: a review and perspective, with remarks on some further crustaceans. Biogerontology 11:643-669.
Vogt G (2011) Marmorkrebs: natural crayfish clone as emerging model for various biological disciplines. J Biosci 36:377-382.
Vogt G, Tolley L, Scholtz G (2004) Life stages and reproductive components of the Marmorkrebs (marbled crayfish), the first parthenogenetic decapod crustacean. J Morphol 261:286-311.
Vogt G, Huber M, Thiemann M, van den Boogaart G, Schmitz OJ, Schubart CD (2008) Production of different phenotypes from the same genotype in the same environment by developmental variation. J Exp Biol 211:510-523.
Vogt G, Wirkner CS, Richter S (2009) Symmetry variation in the heart-descending artery system of the parthenogenetic marbled crayfish. J Morphol 270:221-226.
Vogt G, Falckenhayn C, Schrimpf A, Schmid K, Hanna K, Panteleit J, Helm M, Schulz R, Lyko F (2015) The marbled crayfish as a paradigm for saltational speciation by autopolyploidy and parthenogenesis in animals. Biology open 4:1583-1594.
Weller C, Hochhaus AM, Wright TM, Jr., Mulloney B (2015) A classic improved: minor tweaks yield major benefits in crayfish slow-flexor preparations. J Undergrad Neurosci Educ 13:A74-80.
Wyttenbach RA, Johnson BR, Hoy RR (2018) Reducing the Cost of Electrophysiology in the Teaching Laboratory. Journal of Undergraduate Neuroscience Education 16:277.
Yazawa R, Watanabe K, Koyama T, Ruangapan L, Tassanakajon A, Hirono I, Aoki T (2005) Development of gene transfer technology for black tiger

shrimp, Penaeus monodon. J Exp Zool A Comp Exp Biol 303:1104-1109.
Zhang Y, Oliva R, Gisselmann G, Hatt H, Guckenheimer J, Harris-Warrick RM (2003) Overexpression of a hyperpolarization-activated cation current (Ih) channel gene modifies the firing activity of identified motor neurons in a small neural network. J Neurosci 23:9059-9067.
Zieger E, Braunig P, Harzsch S (2013) A developmental study of serotonin-immunoreactive neurons in the embryonic brain of the marbled crayfish and the migratory locust: evidence for a homologous protocerebral group of neurons. Arthropod Structure & Development 42:507-520.

Received June 23, 2020, revised September 1, 2000; accepted September 21, 2020.

Address correspondence to: Margaret DeMaegd, Illinois State University, School of Biological Sciences, Campus Box 4120, Normal, IL. Email: mdemaeg@ilstu.edu

www.funjournal.org

APPENDIX

Name	Accession	Length (bp)	State of sequence
Artemia parthenogenetica autophagy protein 8 mRNA	KP317126.1	552	complete cds
Caligus clemensi clone ccle-evs-506-228 Microtubule-associated proteins 1A/1B light chain 3A precursor putative mRNA	BT080304.1	1479	complete cds
Caligus clemensi clone ccle-evs-513-169 Gamma-aminobutyric acid receptor-associated protein putative mRNA	BT080241.1	706	complete cds
Caligus clemensi clone ccle-evs-517-116 Gamma-aminobutyric acid receptor-associated protein putative mRNA	BT080712.1	671	complete cds
Caligus rogercresseyi clone crog-evp-507-052 Microtubule-associated proteins 1A/1B light chain 3A precursor putative mRNA	BT077149.1	692	complete cds
Caligus rogercresseyi clone crog-evp-507-179 Gamma-aminobutyric acid receptor-associated protein putative mRNA	BT075883.1	1235	complete cds
Caligus rogercresseyi clone crog-evp-510-035 Microtubule-associated proteins 1A/1B light chain 3A precursor putative mRNA	BT075882.1	736	complete cds
Caligus rogercresseyi clone crog-evp-520-148 Microtubule-associated proteins 1A/1B light chain 3A precursor putative mRNA	BT077123.1	1373	complete cds
Cancer borealis GABA receptor alpha-like protein mRNA	KU986873.1	1785	partial cds
Cancer borealis GABA receptor beta type 1 mRNA	KU986868.1	2598	complete cds
Cancer borealis GABA receptor beta type 2 mRNA	KU986869.1	2534	partial cds
Cancer borealis GABA receptor beta type 3 mRNA	KU986870.1	3236	partial cds
Cancer borealis GABA receptor LCCH3-like protein mRNA	KU986871.1	1446	complete cds
Cancer borealis GABA receptor RDL-like protein mRNA	KU986872.1	1122	partial cds
Cherax quadricarinatus clone V12-216 gabarap protein mRNA	JF284569.2	679	complete cds
Eriocheir sinensis gamma-aminobutyric acid receptor associated protein mRNA	HM149771.1	457	complete cds
Homarus americanus GABA receptor alpha-like protein mRNA	KU986877.1	1764	complete cds
Homarus americanus GABA receptor beta type 1 mRNA	KU986874.1	2523	complete cds
Homarus americanus GABA receptor beta type 2 mRNA	KU986875.1	2540	partial cds
Homarus americanus GABA receptor LCCH3-like protein mRNA	KU986878.1	1446	complete cds
Homarus americanus GABA receptor RDL-like protein mRNA	KU986876.1	996	complete cds
Homarus americanus ionotropic GABA receptor beta subunit 1a mRNA	AY098945.1	2645	complete cds
Homarus americanus ionotropic GABA receptor beta subunit 1b mRNA; alternatively spliced	AY098943.1	2753	complete cds
Lepeophtheirus salmonis clone lsal-evj-015-183 Gamma-aminobutyric acid receptor-associated protein mRNA	BT120928.1	907	complete cds
Lepeophtheirus salmonis clone lsal-evj-025-321 Gamma-aminobutyric acid receptor-associated protein mRNA	BT120613.1	891	complete cds
Lepeophtheirus salmonis Pacific form clone lsal-evj-520-276 Gamma-aminobutyric acid receptor-associated putative protein mRNA	BT077486.1	885	complete cds
Lepeophtheirus salmonis Pacific form clone lsal-evj-522-314 Gamma-aminobutyric acid receptor-associated protein putative mRNA	BT077895.1	909	complete cds
Lepeophtheirus salmonis Pacific form clone lsal-evj-544-239 Gamma-aminobutyric acid receptor-associated protein putative mRNA	BT078746.1	909	complete cds
Litopenaeus vannamei autophagy-related protein 8 mRNA	JQ410230.1	360	complete cds
Panulirus argus ionotropic GABA receptor (GABAR) mRNA	GQ252690.1	110	partial cds
Penaeus monodon autophagy-related protein 8 mRNA	JQ410231.1	360	complete cds
Penaeus vannamei breed Kehai No.1 LVANscaffold_474, whole genome shotgun sequence	QCYY01000474	566471	whole genome shotgun
PREDICTED: *Eurytemora affinis* microtubule-associated proteins 1A/1B light chain 3A-like (LOC111715769), mRNA	XM_023491140	685	complete cds
Procambarus clarkii GABAA receptor subunit mRNA	KM115031.1	1906	complete cds
Procambarus clarkii gamma-aminobutyric acid receptor-associated protein (gabarap) mRNA	MG910480.1	357	partial cds

Appendix Table 1. Results of Nucleotide NCBI search, (GABA receptor) "crustaceans"[porgn:__txid6657]. State of the sequence is given in terms of the coding sequences (cds). Accession numbers can be used to find detailed results of each transcript and associated FASTA files. The highlighted transcripts signify genes of interest.

A

Neurophysiological question: Which type of GABA receptors contribute to synaptic inhibition in the marbled crayfish?

B

Gene of interest Abbrev. Name	NCBI title	Nucleotide sequence
Cancer borealis GABA receptor LCCH3-like	*Cancer borealis* GABA receptor LCCH3-like protein mRNA, complete cds	ATGAGGTGG
Procambarus clarkii GABAA receptor subunit	*Procambarus clarkii* GABAA receptor subunit mRNA, complete cds	CGGGAGCA

C

Homologous marbled crayfish transcript								
Gene of interest	Number	Score (bits)	E value	Transcript sequence	Conserved domain	Accession	Interval	E value
Cancer borealis GABA receptor LCCH3-like	18728	846	0	AAAAACAG	Ligand-gated ion channel super family	TIGR00860	943-2361	1.28E-167
Procambarus clarkii GABAA receptor subunit	2556	2839	0	GGGAGTCT	Ligand-gated ion channel super family	cl36748	489-1589	1.62E-94

D

Genome Scaffold								
Aligned sequence	Scaffold number	Score	E value	Scaffold sequence	Conserved domain	Accession	Interval	E value
Transcript 18728	10872	1552	0	CAGGAATT	LGIC ECD super family	cl28912	2435-2590	1.22E-31
Cancer borealis GABA receptor LCCH3-like	389050	216	2E-53	GGCGTAC	LGIC ECD super family	cl28912		

Appendix Table 2. Example tables for recording and organizing data associated with gene identification which can be used to assess students' understanding. Organizing tables into distinct categories can support accurate containment of information. Including space for the neurophysiological question *(A)*, and tables for the genes of interest *(B)*, the homologous marbled crayfish transcripts *(C)*, and the homologous marbled crayfish scaffolds *(D)*. A complete version of this layout is available as a Google Sheet (https://docs.google.com/spreadsheets/d/1tcryRp-NXFyOuL00auSDebxgJPLtJME2JyecR5a29rk/edit?usp=sharing).

***Cancer borealis* GABA receptor LCCH3-like protein**		
Transcripts producing significant alignments	*Score (bits)*	*E value*
18728	846	0
12698	58	2.00E-07
14362	40	0.039
14447	38	0.15
19560	36	0.61
11501	36	0.61
10839	36	0.61
20118	34	2.4
17973	34	2.4
17156	34	2.4
17049	34	2.4
16855	34	2.4
16352	34	2.4
15747	34	2.4
14065	34	2.4
13608	34	2.4
9926	34	2.4
8996	34	2.4
22094	32	2.4
20192	32	2.4
19376	32	2.4
19270	32	2.4
17684	32	2.4
17325	32	2.4
17201	32	2.4
17030	32	2.4
16101	32	2.4
15614	32	2.4
14816	32	2.4
14784	32	2.4
14610	32	2.4
14364	32	2.4
14345	32	2.4
14023	32	2.4
13032	32	2.4
11132	32	2.4
10880	32	2.4
10230	32	2.4
10208	32	2.4
9252	32	2.4
8182	32	2.4
6791	32	2.4
6709	32	2.4
6031	32	2.4
4762	32	2.4
3784	32	2.4
2940	32	2.4
2516	32	2.4
2481	32	2.4

***Procambarus clarkii* GABAA receptor subunit**		
Transcripts producing significant alignments	*Score (bits)*	*E value*
2556	2839	0
18728	52	1.00E-05
9841	46	8.00E-04
11518	38	0.2
4719	36	0.81
17809	34	3.2
15562	34	3.2
14640	34	3.2
14481	34	3.2
14282	34	3.2
13809	34	3.2
9070	34	3.2

Appendix Table 3. Marbled crayfish transcripts with significant sequence alignments to *Cancer borealis* GABA receptor LCCH3-like protein (left) or *Procambarus clarkii* GABAA receptor subunit (right).

TR18728 - GABAA receptor LCCH3 subunit

Scaffolds producing significant alignments	*Score (bits)*	*E value*
10872	1552	0
389050	624	1.0E-176
239177	52	0.001
4465	44	0.36
277499	42	1.4
49753	42	1.4
11235	42	1.4
4188	42	1.4
4065	42	1.4
391036	40	5.6
200749	40	5.6
134921	40	5.6
46040	40	5.6
27780	40	5.6
17349	40	5.6
10065	40	5.6
10044	40	5.6

TR2556 - GABAA receptor subunit

Scaffolds producing significant alignments	*Score (bits)*	*E value*
239177	1116	0
10872	52	0.001
385812	42	0.87
76221	42	0.87
71437	40	3.4
8174	40	3.4

Appendix Table 4. Marbled crayfish genome scaffolds with significant sequence alignments to putative GABAA receptor LCCH3 subunit (left) or GABAA receptor subunit (right).