Zebrafish as a Model System for Research and Teaching
2019 *Eastern Biologist* Special Issue 1:47–63

Zebrafish Embryos and Bioinformatics: Useful and Marketable Exercises for Students Enrolled in Upper-Level Undergraduate Courses

Adam Davis[1*], Hong Nguyen[1], and Jo Qian[1]

Abstract - The zebrafish (*Danio rerio*) is a widely used vertebrate model system in several branches of biological research, including molecular biology, developmental biology, and evolutionary biology. The rapid and transparent development of embryos outside of the mother allows for real-time observation of embryonic development of different types of organ systems. However, the types of wet-lab exercises involving zebrafish embryos that undergraduate students can perform during a 2-3 hour laboratory period are limited. Recent advances in bioinformatics applications and the availability of genomic sequence data from zebrafish and other evolutionarily divergent vertebrates, including human, allow students to be actively involved in semester-long molecular, evolutionary, and developmental biology projects that can be performed both in and out of the laboratory. In this paper, we report several exercises that can be used in upper level undergraduate biology courses that include a 2-3 hour weekly laboratory session. These exercises include amino acid and genomic DNA sequence alignment and gene expression analysis of *Hoxa2*, a developmental regulatory gene that is highly characterized in its expression and function. All exercises make use of standard operating procedures for training students on new techniques. The exercises presented will provide several learning outcomes for students, including the identification of conserved protein domains and *cis*-regulatory elements and how mutations to these motifs can lead to evolutionary diversity as well as the development of homeostatic imbalances in humans.

Introduction

The zebrafish (*Danio rerio*) provides an excellent system for allowing students to gain an appreciation on how vertebrate model organisms are utilized in evolutionary, molecular, developmental, toxicological, and biomedical studies (D'Costa and Shepherd 2009, Sarmah et al. 2016, Schmoldt et al. 2009). Several beneficial characteristics of zebrafish allow them to be used by students in wet-lab genetic studies. These include, but are not limited to, high brood volume, all-year embryo accessibility, and transparent embryos that develop rapidly outside of the mother (see Gilbert and Barresi 2016). While these characteristics allow students to visualize real-time development using microscopic techniques, the use of these techniques is generally constrained to a 2-to-3 hour laboratory environment. Thus, the potential number and types of laboratory exercises tend to be limited. This can be unfortunate for students that are enrolled in upper-level undergraduate biology courses but are not actively involved in independent research projects.

Beyond the qualities that make zebrafish an excellent laboratory model, the entire genome as well as smaller regions of genomic DNA of this species, are published over several online databases. The ease of access of protein and genomic DNA sequences from zebrafish and other vertebrate organisms from these databases allows for the generation of several bioinformatics-based student-run projects that are not constrained to a laboratory or classroom environment. The performance of bioinformatics-based projects using zebrafish

[1]Biology Department, University of North Georgia, Oakwood, GA 30566, USA. [*]Corresponding author: adam.davis@ung.edu.

genetic data allows students to gain hands-on, marketable skills that can translate to careers in industrial, government, and academic settings (Cattley and Arthur 2007, Cohen 2003, Ditty et al. 2010, Floriano 2008, Maloney et al. 2010). Further, such studies also provide students the understanding on how and why specific model organisms beyond primates are used for genetic research, including pharmaceutical drug discovery (Caroll et al. 2003).

Here, we outline several exercises, both bioinformatics-based and wet-lab-based, that can be performed over the course of a semester-long, student-run project in upper level undergraduate genetics-based biology courses. These techniques are currently being used in Evolutionary and Developmental biology courses taught at University of North Georgia. Students are guided through these techniques by using standard operating procedures (SOPs). All SOPs contain an objective, a list of relevant terms and their definitions, the procedure in outline format, references, and an assessment with a sign-off page. SOPs aid in the training of new techniques for students and ensure that students perform these techniques in the proper and logical order (Bhattacharya 2015). All SOPs used for the exercises listed below can be located in the SOP Supplemental file (see supplemental File 1, available online at https://eaglehill.us/ebioonline/suppl-files/ebio-022-davis-s1.pdf). All exercises focus on analyzing *Hoxa2*, a homeodomain-containing transcription factor that functions, in part, to pattern many of the cranial nerves and craniofacial cartilages and bones during vertebrate embryonic development. This gene was chosen for these analyses for several reasons: 1) its expression and function are well documented across several vertebrate model systems (Baltzinger et al. 2005; Davenne et al. 1999; Davis et al. 2008; Gavalas et al. 1997; Gendron-Maguire et al. 1993; Grammatopoulus et al. 2000; Hunter and Prince 2002; Le Pabic et al. 2007, 2010; Pasqualetti et al. 2000; Prince and Lumsden 1994; Rijli et al. 1993; Scemama et al. 2006), 2) its protein structure, as well as the domains that enable its function, have been identified (Chang et al. 1996, LaRonde-LeBlanc and Wolberger 2003, Piper et al. 1999), and 3) the genomic *cis*-regulatory elements (CREs) that direct when and where this gene is expressed for the proper development of vertebrate head anatomy have been identified using the mouse, chicken, and several fish model systems (Amin et al. 2015; Davis et al. 2016; Frasch et al. 1995; Lampe et al. 2008; Maconochie et al. 1999, 2001; McEllin et al. 2016; Nonchev et al. 1996; Parker et al. 2014; Tümpel et al. 2002, 2006, 2007, 2008, 2009). Beyond using *Hoxa2* as the candidate gene, we show the importance of comparing human *Hoxa2* amino acid and genomic DNA sequences to those of other closely related primates as well as known genetic vertebrate model systems, including zebrafish. Zebrafish, human, mouse, and chicken share a most recent ancestor that lived roughly 400 million years ago (mya) (Benton and Donoghue 2007, Broughton et al. 2013), and such evolutionary history will aid in the identification of conserved and functional sequences.

Representative Exercises Used for Semester-Long Projects

Exercise #1: The use of zebrafish and other vertebrate model organisms in the identification of functional *Hoxa2* protein domains

Bioinformatics analyses of *Hoxa2* amino acid sequence alignments between distantly related vertebrates allow students to obtain hands-on experience in identifying functional domains of proteins that are integral to human physiology and development (Tenorio 2014). To fully understand why distantly related vertebrates are necessary in such analyses, students perform two separate alignments, with each including the human *Hoxa2* amino acid sequence. The first alignment compares the human *Hoxa2* amino acid sequence with that of other closely related primate species, including chimpanzee (*Pan*

troglodytes), western lowland gorilla (*Gorilla gorilla*), and sumatran orangutan (*Pongo abelii*). The other analysis compares human with orthologous sequences of zebrafish, chicken (*Gallus gallus*), and mouse (*Mus musculus*).

For this exercise, students are first trained on a SOP involving the extraction of species-specific *Hoxa2* amino acid sequences from the National Center for Biotechnology Information (NCBI) website (see BIO-001 SOP in Supplemental file). All sequences are extracted in FASTA format by copying the sequences and pasting them to a Microsoft Word file. The NCBI accession numbers for the amino acid sequences are as follows: Human (accession number: NP_006726), Chimpanzee (XP_527697), Gorilla (XP_004045263), Orangutan (XP_002818153), Mouse (NP_034581), Chicken (NP_990481), and Zebrafish (NP_571181).

Once the Microsoft Word file containing the extracted *Hoxa2* amino acid sequences is complete, the sequences must be aligned and color-coded. Students are trained using the same SOP (BIO-001) to perform the alignment, color-coding processes, and transfer of coded sequences into Microsoft PowerPoint. This SOP also covers the generation of percent identity matrices for students to observe the amino acid sequence divergence between evolutionarily divergent species. Alignments are generated by using the Clustal Omega software program over the EMBL-EBI web site (https://www.ebi.ac.uk/) (Larkin et al. 2007). Once alignments are generated, they are copied over to Microsoft word for further formatting and color-coding of amino acids (Fig. 1). Amino acid columns that show 100% sequence identity are color-coded yellow. Amino acids within columns that show 75% sequence identity are color-coded blue. Amino acids within columns that show 50% sequence identity or less are left uncolored. Gaps, which represent insertion/deletion (indel) mutations are left uncolored. Color-coded amino acid alignments are then copied and pasted as images into Microsoft PowerPoint so that experimentally determined functional domains can be labeled using this software (Fig. 1). Percent identity matrices are also computed from aligned data over the EMBL-EBI website (Table 1).

Once alignments are generated, formatted and color-coded, students observe noticeable differences between the two data sets. *Hoxa2* protein functional domains will not be readily observable in the primate-specific alignment (Fig. 1). This is due to the high sequence identity between primate sequences, wherein almost all amino acids are identical between primates (Table 1). The presence of high sequence identity is due to these primates sharing a most recent common ancestor that only lived roughly 20 mya, which obscures students' abilities to locate functional domains (Steiper and Young 2006). By contrast, the reduced sequence identity

Table 1. Percent similarity of *Hoxa2* amino acid sequences between human and three other primates and between Human and three model vertebrate organism.

Model	Human	Chimpanzee	Gorilla	Orangutan
Human	-	100	99.20	98.66
Chimpanzee	100	-	99.20	98.66
Gorilla	99.20	99.20	-	98.40
Orangutan	98.66	98.66	98.40	-

Model	Human	Mouse	Chicken	Zebrafish
Human	-	96.49	87.40	72.10
Mouse	96.49	-	87.03	71.67
Chicken	87.40	87.03	-	71.35
Zebrafish	72.10	71.67	71.35	-

Figure 1. Comparative amino acid sequence analysis of the *Hoxa2* protein between orangutan, gorilla, chimpanzee, and human (left alignment) and zebrafish, chicken, mouse, and human (right alignment). Amino acids colored in yellow correspond to complete conservation at particular sites across all sequences examined. Amino acids colored in blue represent the majority of the sequences containing specific amino acids at specific sites. Black boxed regions in the right alignment correspond to functional *Hoxa2* domains.

of *Hoxa2* protein sequences between human and the three model vertebrate organisms, zebrafish, chicken, and mouse, helps to reveal several conserved *Hoxa2* protein domains, including the hexapeptide and homeodomain motifs (Fig. 1). To identify these motifs, students are trained on a SOP that covers the use of PROSITE, an online database of protein domains, families, and functional sites (Sigrist et al. 2002) (see BIO-002 in Supplemental file). The human *Hoxa2* amino acid sequence can be used in the PROSITE software for students to locate these domains. Instructors must provide literature to students to show that the hexapeptide functions to bind with Pbx proteins, and the homeodomain, along with the Pbx DNA-binding domain, functions to bind to Hox/Pbx CREs within genomic DNA to regulate the spatial and temporal expression patterns of downstream genes (Chang et al. 1996, LaRonde-LeBlanc and Wolberger 2003, Piper et al. 1999). Finally, students are trained using the same SOP (BIO-002) to label the hexapeptide and homeodomain using transparent rectangles and textboxes in Microsoft PowerPoint (see Fig. 1).

Table 2 provides a suggested sequence of steps to follow for performing Exercise #1. This exercise helps to highlight several learning outcomes for students. First, it suggests functional significance of the hexapeptide and homeodomain motifs, since they have persisted without modification in amino acid sequence in vertebrates for roughly 400 million years. These analyses suggest that *Hoxa2* performs similar cellular functions within evolutionarily divergent species. Further, these analyses show that, in comparison to amino acid sequences surrounding the hexapeptide and homeodomain, the DNA sequences that give rise to these domains may be experiencing increased evolutionary constraint due to increased purifying selection. Nonsynonymous mutations giving rise to amino acid changes within the homeodomain or hexapeptide sequences would potentially abrogate the function of these domains and the entire *Hoxa2* protein, thus reducing the fitness of an individual containing these mutations. Second, this exercise offers evidence that humans share relatively recent common ancestors with other primates, since their entire *Hoxa2* amino acid sequences are shown to be nearly identical to each other (Fig. 1).

Exercise #2: The use of zebrafish and other vertebrate model organisms in the identification of functional *cis*-regulatory elements that direct *Hoxa2* gene expression

Cis-regulatory elements (CREs) are genomic DNA sequences that function as binding sites for transcription factor proteins and direct the expression of their associated genes in specific developmental and anatomical compartments (Davidson 2006). As in finding con-

Table 2. Suggested sequence of steps to be performed for Exercise #1.

Step 1	Train students using the first SOP, BIO-001 (see Supplemental file), which covers: 1) Extraction of species-specific amino acid sequences in FASTA format from NCBI. 2) Generation of amino acid sequence alignments using Clustal Omega software. 3) Color-coding of amino acid sequence alignments in Microsoft Word 4) Transfer of color-coded sequences into Microsoft PowerPoint 5) Generation of Percent Identity Matrices of aligned amino acid sequence alignments from Clustal Omega. Students turn in assigned work for grading.
Step 2	Train students using the second SOP, BIO-002 (see Supplemental file), which covers: 1) Identification of conserved protein domains using PROSITE software 2) Labeling of hexapeptide and homeodomain protein domains using Microsoft PowerPoint Students turn in assigned work for grading.

served protein domains in amino acid sequence alignments, the location of CREs requires the alignment of genomic DNA sequences from distantly related organisms. Amazingly, *Hoxa2* shows a conserved expression pattern within rhombomeres (r) 2-5 of the hindbrain and the pharyngeal arches (PAs), or transient embryonic head compartments that give rise to the cranial nerves and craniofacial skeletal elements, respectively. Many of the CREs that direct *Hoxa2* in these rhombomeres and pharyngeal arches have been mapped in the mouse model system and shown to be conserved across vertebrates, from fish to mammals (Amin et al. 2015; Davis et al. 2016; Frasch et al. 1995; Maconochie et al. 1999, 2001; McEllin et al. 2016; Nonchev et al. 1996; Parker et al. 2014; Tümpel et al. 2002, 2006, 2007, 2008, 2009). The CREs that direct *Hoxa2* gene expression in r3 and r5 and the PAs are located in noncoding DNA upstream of *Hoxa2* (Amin et al. 2015; Davis et al. 2016; Frasch et al. 1995; Maconochie et al. 1999 and 2001; McEllin et al. 2016; Nonchev et al. 1996; Parker et al. 2014; Tümpel et al. 2002, 2006, 2009). The CREs that direct *Hoxa2* expression in r4 are located within the *Hoxa2* intron (Tümpel et al. 2006, 2007, 2009). The CREs that direct *Hoxa2* expression in r2 are located within exon 2 of *Hoxa2* (Tümpel et al. 2006, 2008, 2009). Students can visualize many of these CREs both in a global scale using mVista analysis and in a more detailed and localized scale using Clustal sequence alignment. Although the CREs that direct *Hoxa2* expression in r2 have been identified and functionally tested, they are located within coding DNA and are not as easily discernible to students as those specific to r3-r5 and the PAs. Therefore, we suggest that the instructor not cover the location of these CREs for this exercise. However, we have listed the sequence coordinates for the genomic sequences that contain several of these CREs below if the instructor chooses to incorporate these elements into this exercise.

For this exercise, students are first trained on a SOP involving the extraction of species-specific *Hoxa2* genomic DNA sequences from the NCBI website and the generation of sequence annotation files (see BIO-003 in Supplemental file). Genomic sequences correspond to the exons, intron, and 3000 bp upstream of the ATG start site of *Hoxa2.* Genomic sequences are extracted from all seven organisms used for the amino acid exercise shown above. Students must obtain reverse complement sequences for most genomic sequences from the NCBI database in order to produce a mVista plot with all genomic DNA sequences in the same orientation. Each sequence is in FASTA format and saved as a separate simple text file for each organism. Sequence annotation files are constructed to show the sequence coordinates of the coding regions for each sequence. Individual sequence annotation files corresponding to each genomic DNA sequence are also saved in FASTA format and as simple text files. NCBI accession numbers and sequence coordinates for genomic DNA containing *Hoxa2* and 3000 bp upstream of *Hoxa2* and the coordinates for the exons used in sequence annotation files are listed in Table 3. These coordinates are incorporated into the BIO-003 SOP for training students.

Once all species-specific sequence files and their associated sequence annotation files are produced, students are trained using the same SOP (BIO-003) on the use of the mVista program (http://genome.lbl.gov/vista/index.shtml) and the generation of graphical figures of genomic sequence comparisons using Microsoft PowerPoint. mVista is used to display global sequence alignments from multiple species (Frazer et al. 2004, Mayor et al. 2000). Outputs of sequence alignments allow students to visualize regions of high sequence identity over long stretches of genomic sequence in graphical format. Further, this program allows for the identification of conserved regions that have experienced inversion or translocation mutations. Students should construct a mVista plot using all seven species listed above. The LAGAN option should be used for this exercise, since it performs progressive pairwise

sequence alignments and the genomic region examined does not exhibit rearrangements between species (Brudno et al. 2003). When mVista plots are generated, students use the human genomic DNA sequence as the reference sequence. Students observe differences in sequence identity across several regions of genomic DNA when comparing the human genomic DNA to other vertebrates (Fig. 2). Red and blue peaks correspond to conservation in noncoding and coding DNA, respectively. Students observe that the orangutan, gorilla, and chimpanzee sequences show almost complete sequence identity with human (Fig. 2). Further, mouse shows greater sequence identity with human than either zebrafish or chicken. Interestingly, zebrafish only shows high sequence identity with human in the *Hoxa2* exons, the proximal promoter region, a sequence motif upstream of the proximal promoter, and an intronic region (Fig. 2). The presence of conserved domains suggests increased evolutionary constraint in these regions. While the exons and proximal promoter region are expected to show high conservation, since they are involved in coding for the *Hoxa2* amino acid sequence and transcription of this gene, respectively, the conservation in sequence identity of the upstream and intronic regions suggest that these regions function in the regulation of *Hoxa2* gene expression. Students convert the generated mVista plot images using the human genomic DNA sequence as a reference sequence to a figure that is modified using Microsoft

Table 3. NCBI accession numbers and genomic DNA sequence coordinates for *Hoxa2* of all organisms used for the mVista analysis. Coordinates for start and end positions of exons 1 and 2 for all *Hoxa2* genes are listed. Sequences that require the reverse complement to obtain the 5' to 3' orientation are marked with an asterisk (*) in the Model column.

Model	Accession	Start	End	Exon 1 Start	Exon 1 End	Exon 2 Start	Exon 2 End
Zebrafish*	AL645795	56638	61314	3001	3361	3947	4677
Chicken*	NC_006089	32582800	32587527	3001	3369	3981	4728
Mouse*	CH466597	4372386	4377144	3001	3379	4018	4759
Orangutan	NC_036910	46400034	46404802	3001	3385	4030	4769
Gorilla*	NC_018431	27155539	27160306	3001	3388	4029	4768
Chimpanzee*	NC_036886	27217325	27222099	3001	3391	4036	4775
Human*	AC004079	83121	87895	3001	3391	4034	4775

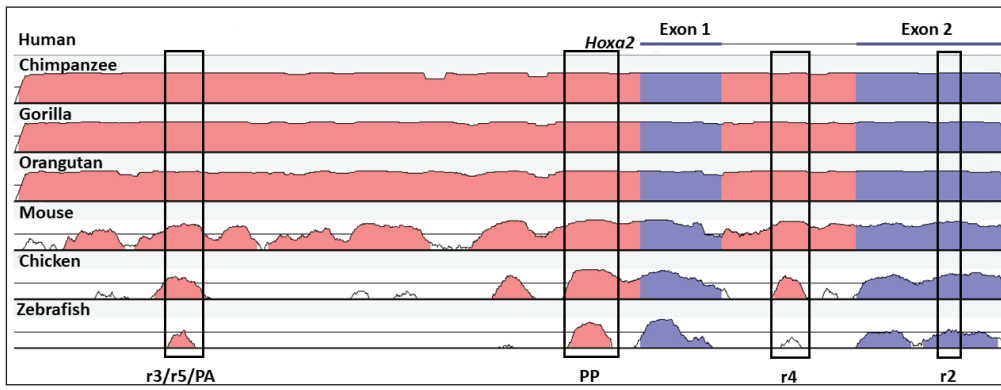

Figure 2. mVista analysis of orthologous genomic DNAs corresponding to 3000 bp upstream of *Hoxa2* and *Hoxa2* itself. Human is used as the reference sequence to which all other genomic DNAs are compared, including zebrafish, chicken, mouse, orangutan, gorilla, and chimpanzee. Red peaks correspond to noncoding DNA. Blue peaks correspond to Exons 1 and 2 of *Hoxa2* (labeled). Black boxed regions correspond to genomic DNA regions responsible for directing the expression of *Hoxa2*. r3/r5/PAs, genomic DNA region that contains the CREs that direct *Hoxa2* expression in r3, r5, and the PAs; PP, proximal promoter; r4, genomic DNA region that directs *Hoxa2* in r4; r2, genomic DNA region that directs *Hoxa2* in r2.

PowerPoint. mVista images are copied through the print screen option on the keyboard and pasted as images in Microsoft PowerPoint. Transparent rectangles and textboxes are used to label the regions that direct *Hoxa2* gene expression in r3, r5, and the PAs, r4, and the region corresponding to the proximal promoter. Students can view a similar published image in Davis et al. (2008) to be used as a guide in labeling the r3/r5/PAs and r4-specific regions.

Once mVista images are modified, further analysis requiring Clustal alignment of the upstream enhancer region and intronic region allows students to reveal the CREs that direct *Hoxa2* expression in r3/r5/PAs and r4, respectively. Students are trained on SOPs involving the extraction of species-specific *Hoxa2* genomic sequences containing functionally tested regulatory elements from the NCBI website (see BIO-004 and BIO-005 SOPs for the r3/r5/PA and r4 CREs, respectively, in the Supplemental file). Students must obtain reverse complement sequences for most genomic sequences from the NCBI database in order to produce correct sequence alignments using the Clustal software program. All sequences are extracted in FASTA format by copying the sequences and pasting them to two separate Microsoft Word files: one specific for the r3/r5/PA CREs and another for the r4 CREs. Sequence coordinates for genomic DNA sequences containing the CREs of interest are listed in Table 4. Coordinates for the r2 CREs are also listed in Table 4 if instructors would like to show students that CREs can be embedded within coding DNA sequences. Further, sequences requiring the acquirement of reverse complements are listed in Table 4.

Once the Microsoft Word files have been generated the sequences pertaining to each region containing specific CREs must be aligned and color-coded. Students are trained using the same SOPs (BIO-004 and BIO-005) but that cover (1) the alignment of sequences using Clustal Omega software, (2) color-coding processes similar to those used for the amino acid alignments, and (3) the transfer of coded sequences into Microsoft PowerPoint. Further, as mentioned above, two alignments using human *Hoxa2* genomic DNA should be generated for each of the genomic DNA regions containing CREs involved in rhomobomere and PA expression. Human should be compared to the three other primates in one set of alignments and against zebrafish, chicken and mouse in the other set.

Once genomic DNA sequence alignments are generated, formatted and color-coded, students observe noticeable differences between the two data sets. As in the amino acid sequence alignments shown in Exercise #1 above, functional CREs are more readily observable in the alignments involving human, zebrafish, chicken, and mouse than with human and the other primates (Figs. 3 and 4). To identify the CREs, students must be provided with lit-

Table 4. Genomic DNA sequence coordinates for the localized alignment of CREs contributing to *Hoxa2* expression in r2-r5 and the PAs. NCBI Accession numbers for all organisms are listed in Table 3. Sequences that require the reverse complement to obtain the 5' to 3' orientation are marked with an asterisk (*) in the Model column.

	r3/r5/PAs alignment coordinates		r4 alignment coordinates		r2 alignment coordinates	
Model	Start	End	Start	End	Start	End
Zebrafish*	59409	59802	57649	57741	56827	56886
Chicken*	32586583	32586983	32583964	32584047	32582978	32583051
Mouse*	4376330	4376657	4373429	4373513	4372578	4372637
Orangutan	46400521	46400872	46403675	46403759	46404551	46404610
Gorilla*	27159477	27159825	27156581	27156665	27155731	27155790
Chimpanzee*	27221268	27221616	27218368	27218452	27217517	27217576
Human*	87067	87415	84164	84248	83313	83372

erature that identifies the specific CREs. For instance, Krox20 and Sox transcription factor binding sites are involved in directing *Hoxa2* expression in r3 and r5 (McEllin et al. 2016) and the Hox/Pbx and Prep/Meis sites direct *Hoxa2* in the PAs in the upstream enhancer region (Davis et al. 2016) (Fig. 3). In the intronic DNA, a Prep/Meis and several Hox/Pbx CREs function in tandem to direct *Hoxa2* expression in r4 (Fig. 4A) (Tümpel et al. 2007). If the instructor wishes to include the CREs responsible for directing *Hoxa2* expression in r2, several Sox CREs within exon 2 of *Hoxa2* function to direct this gene's expression in this embryonic domain (Fig. 4B) (Tümpel et al. 2008). The literature cited above show sequence alignment figures with the CREs of interest labeled. Finally, students are trained using the same SOPs (BIO-004 and BIO-005) to label the CREs mentioned above using transparent rectangles and textboxes in Microsoft PowerPoint.

Table 5 provides a suggested sequence of steps to follow for performing Exercise #2. As in the amino acid sequence alignments, these analyses help to highlight several learning outcomes for students. First, they show that, in comparison to surrounding sequences, the CREs involved in *Hoxa2* gene expression may be experiencing increased evolutionary constraint due to increased purifying selection. Mutations to these sequences could potentially disrupt their binding to specific transcription factors and the overall expression of *Hoxa2* from the rhombomeres and PAs. Such mutations could lead to altered genetic regulatory networks within the embryonic compartments in which *Hoxa2* is normally involved, and thus, altered morphogenetic patterning of head-specific organs. Second, these analyses allow students to understand the concept of pleiotropy. Deleterious mutations to DNA sequences giving rise to the homeodomain or hexapeptide could affect genetic regulatory networks in all the rhombomeres and PAs in which *Hoxa2* is expressed. Therefore, a disrupted homeodomain or hexapeptide could lead to the disruption of development of several head-specific organs. Third, the conservation in sequence of CREs that have been functionally mapped in mouse will allow students to hypothesize the expression pattern of *Hoxa2* in zebrafish embryos. Based on the conservation in CRE sequences, students should hypothesize that zebrafish *Hoxa2* will be expressed within r3-5 of the hindbrain and the PAs of zebrafish embryos.

Exercise #3: The use of zebrafish to visualize *Hoxa2* expression during embryonic development

The zebrafish model system provides students with an excellent opportunity to observe the spatial and temporal patterns of gene expression during embryonic development. The transparency of embryos allows students to observe clear gene expression patterns in developing organs of whole-mount embryos. Further, several studies have shown that zebrafish provides an excellent model for visualizing gene expression and performing other laboratory assays in the classroom (D'Costa and Shepherd 2009, Sarmah et al. 2016, Schmoldt et al. 2009). We have developed a SOP that covers the whole-mount *in situ hybridization* (WISH) assay using zebrafish embryos based off published research articles (Davis et al. 2008, Prince et al. 1998) (see BIO-006 in Supplemental file). For this analysis, the *Hoxa2* cDNA clone used for producing riboprobes in WISH analyses are purchased from the Addgene website (https://www.addgene.org/; deposited by V. Prince) (Prince et al. 1998). We have found that this riboprobe works very well in uncovering *Hoxa2* expression patterns in zebrafish embryos. It may be beneficial for the instructor to prepare the experimental antisense and control sense riboprobes and perform the testing of their efficacy on fixed zebrafish embryos prior to the performance of this exercise. Otherwise, SOPs should be developed that detail the manufacturing of sense and antisense riboprobes. Thirty-hour post fertilization (hpf) zebrafish embryos provide an ideal

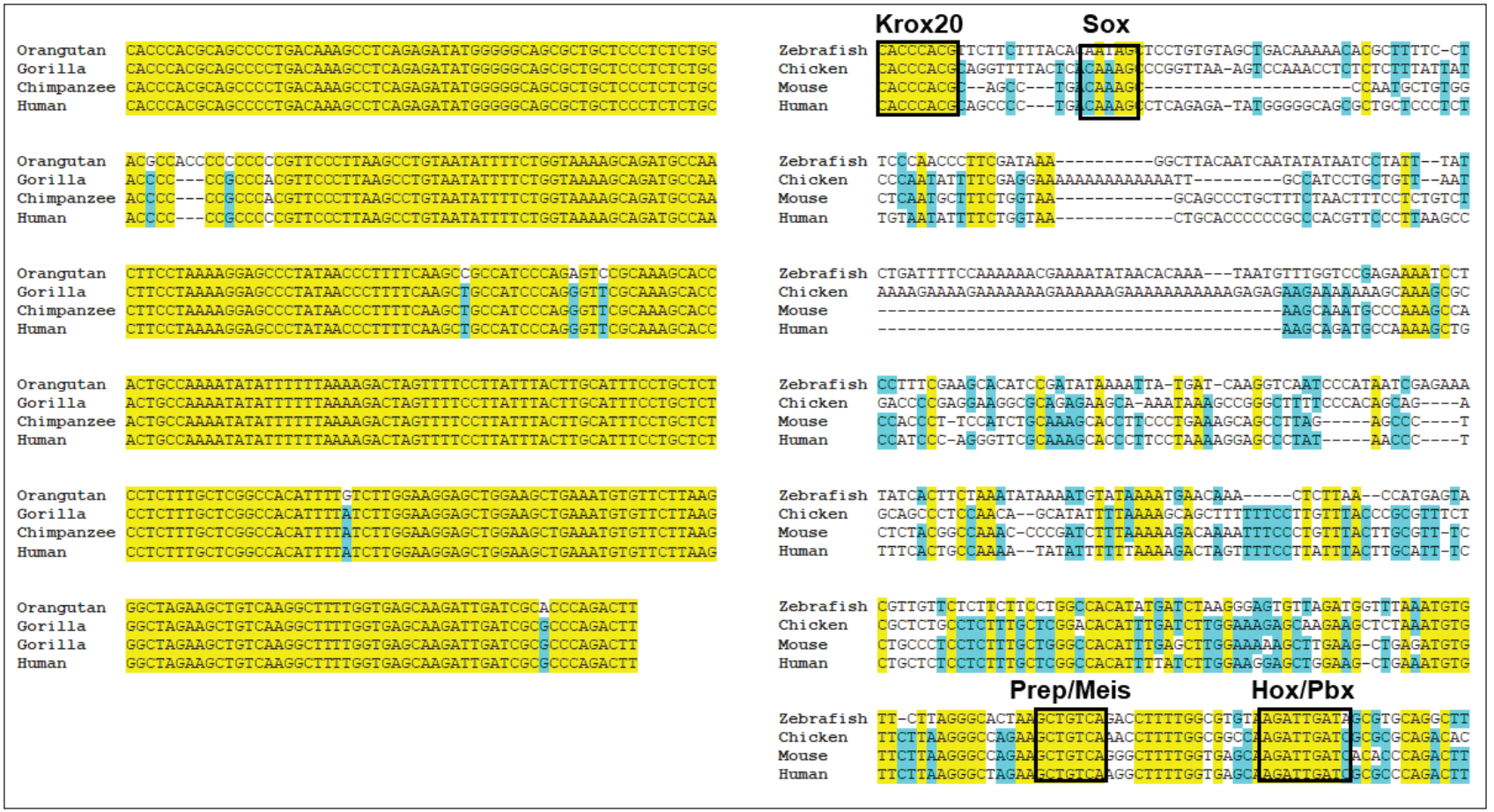

Figure 3. Comparative genomic DNA sequence analysis of the upstream enhancer region containing the CREs that direct *Hoxa2* expression in r3, r5, and the PAs between orangutan, gorilla, chimpanzee, and human (left alignment) and zebrafish, chicken, mouse, and human (right alignment). Base pairs colored in yellow correspond to complete conservation at particular sites across all sequences examined. Base pairs colored in blue represent the majority of the sequences containing specific nucleotides at specific sites. Black boxed regions in the right alignment correspond to CREs that bind specific transcription factors.

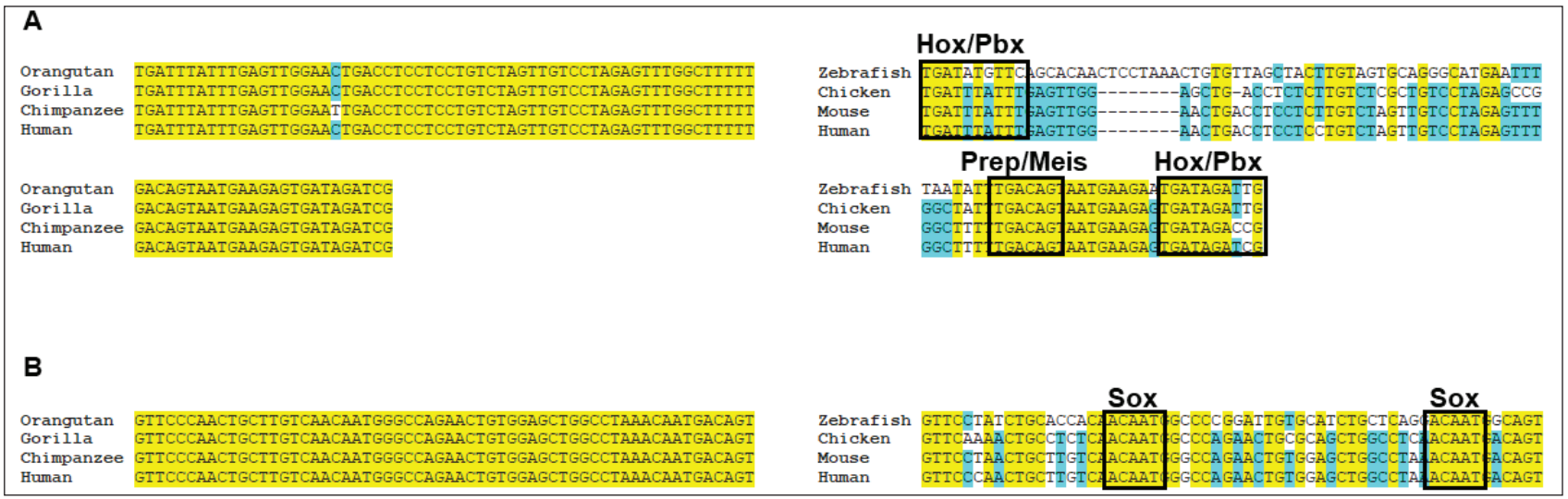

Figure 4. Comparative genomic DNA sequence analyses of the intronic (A) and second exon (B) regions containing the CREs that direct *Hoxa2* expression in r4 and r2, respectively, between orangutan, gorilla, chimpanzee, and human (left alignments) and zebrafish, chicken, mouse, and human (right alignments). Base pairs colored in yellow correspond to complete conservation at particular sites across all sequences examined. Base pairs colored in blue represent the majority of the sequences containing specific nucleotides at specific sites. Black boxed regions in the right alignments correspond to CREs that bind specific transcription factors.

stage for visualizing *Hoxa2* gene expression. Both the rhombomeres and pharyngeal arches are well defined at this stage (Kimmel et al. 1995). While WISH assays are usually performed over a three-day period, several of the stages of these assays allow for long-term storage of embryos at 4 ˚C or -20 ˚C. Therefore, WISH analyses can be performed over the span of several weeks for courses that have just one 2-3 hour lab per week. Students are trained on the SOP that covers the WISH assay (BIO-006). The SOP covers specific points at which embryos can be preserved at 4 ˚C or -20 ˚C.

Once WISH analyses are completed, students must mount embryos in a lateral orientation on microscope slides and record images using microscope visualization software programs. Lateral orientations show both rhombomeres and pharyngeal arches in the same image (Fig. 5). It may be best for the instructor to demonstrate the lateral mounting of embryos on microscope slides and digital photography without the aid of SOPs. The robustness of gene expression assays may differ between individual students or groups of students. Further, teaching laboratories are generally designed to have just one microscope with an accompanying camera system linked to a screen to which the entire class can visualize. We have found that students enjoy participating vocally in the positioning of embryos for the most informative views when regarding the rhombomeres and pharyngeal arches. Once images are agreed upon by the class and saved, they can be further processed in Adobe Photoshop by the instructor and distrib-

Table 5: Suggested sequence of steps to be performed for Exercise #2.

Step 1	Train students using the third SOP, BIO-003 (see Supplemental file), which covers: 1) Extraction of species-specific genomic DNA sequences in FASTA format from NCBI. 2) Generation of annotation files specific to each sequence. 3) Generation of graphical global sequence alignments between all seven species using the species-specific genomic DNA sequence and annotation files in mVISTA software. 4) Modification of graphical sequence alignment image in Microsoft PowerPoint. Students turn in assigned work for grading.
Step 2	Train students using the fourth SOP, BIO-004 (see Supplemental file), which covers: 1) Extraction of species-specific genomic DNA sequences in FASTA format from regions that direct *Hoxa2* expression in r3/r5/PAs. 2) Alignment of r3/r5/PA-specific genomic DNA sequences using Clustal Omega. 3) Color-coding DNA sequence alignments using Microsoft Word. 4) Transfer of color-coded DNA sequence alignments to Microsoft PowerPoint for figure modification. 5) Identification and labeling of CREs that direct *Hoxa2* in r3/r5/PAs using scientific literature and Microsoft PowerPoint. Students turn in assigned work for grading.
Step 3	Train students using the fifth SOP, BIO-005 (see Supplemental file), which covers: 1)Extraction of species-specific genomic DNA sequences in FASTA format from regions that direct *Hoxa2* expression in r4. 2) Alignment of r4-specific genomic DNA sequences using Clustal Omega 3) Color-coding DNA sequence alignments using Microsoft Word. 4) Transfer of color-coded DNA sequence alignments to Microsoft PowerPoint for figure modification. 5) Identification and labeling of CREs that direct *Hoxa2* in r4 using scientific literature and Microsoft PowerPoint. Students turn in assigned work for grading.

uted to the students for the labeling of relevant rhombomeres and pharyngeal arches using Microsoft PowerPoint. This may also be performed without the aid of a SOP, since images of embryos will change from semester to semester. Students should label the PAs and r2-r5 (Fig. 5). Students may also want to label r1 and r6, or rhombomeres that are just anterior and posterior to the *Hoxa2*-expressing rhombomeres, respectively, to show the contrast in *Hoxa2* gene expression between different embryonic modules (Fig. 5).

Table 6 provides a suggested sequence of steps to follow for performing Exercise #3. Students will obtain several learning outcomes from the WISH assays. First, they will observe that *Hoxa2* of zebrafish shows a conserved expression pattern during embryonic development with other vertebrates, specifically in r2-r5 and the PAs. This conserved expression pattern may be due to the sequence conservation of CREs directing *Hoxa2* in these embryonic compartments. Second, the conservation of amino acid sequence in the homeodomain and hexapeptide allows students to understand that *Hoxa2* functions in a conserved manner in these embryonic compartments throughout the vertebrates, from fish to mammals. Visualization of expression patterns further help students understand the concept of pleiotropy, such that mutations to these domains can lead to the disruption of proper rhombomere and pharyngeal arch development and the organs derived from these compartments.

Assessment of Learning

Students are assessed of their learning with the projects mentioned above using several strategies. First, students are assessed with assignments for each step of the exercises shown above. For instance, assignments for Exercise #1 may include turning in Microsoft Word files containing the sequences for alignment in FASTA format, Microsoft Word

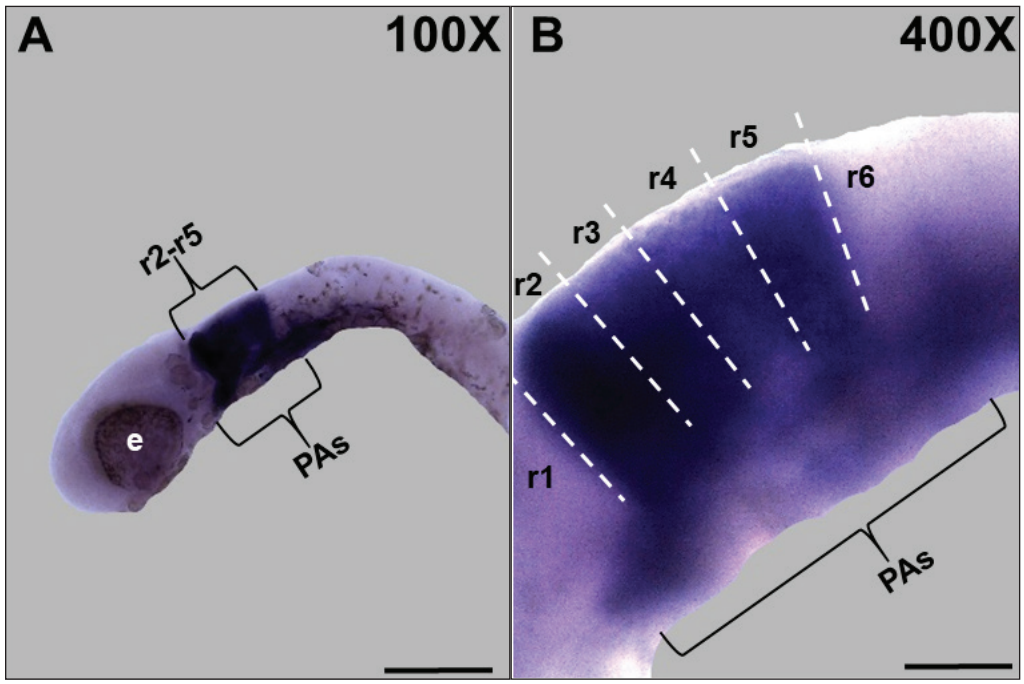

Figure 5. Whole-mount *in situ* hybridization analysis of *Hoxa2* expression in 30 hpf zebrafish embryo. Embryo is shown at 100X (A) and 400X (B) magnification. Embryo is mounted with its anterior side facing left and its lateral side facing the reader. Dashed lines represent rhombomere boundaries. e, eye; PAs, pharyngeal arches; r, rhombomere. Scale bars equal 0.1 mm.

files containing sequence alignments, Microsoft Word files containing properly color-coded sequences, and Microsoft PowerPoint files containing labeled hexapeptide and homeodomain regions. Although students are trained on SOPs, their ability to follow directions is a necessity for their career paths, no matter the path. Second, figures produced by students can be used in lecture to support the teaching of complex concepts, such as protein structure and function, regulation of gene expression, and variation of selective constraint in genomic DNA sequences. The use of smart phone-based application software packages in lecture will allow instructors to assess students' understanding of the exercises (Wash and Freeman 2013). Finally, figures produced by students can be used in lecture and/or laboratory examinations. Each of the figures produced by students can lend for several questions that can be used in examinations.

At the end of the courses in which these SOPs and exercises are used, students are asked to provide anonymous comments on their impressions of using SOPs and performing bioinformatics. The most common answer has been that students enjoy using SOPs because they provide real-life experiences and are applicable to the world outside of academia. Further, students generally enjoy learning marketable bioinformatics skills. Students are also asked to provide feedback on how to improve the exercises. By using this feedback, we have expanded the exercises to include SOPs that are used to train students on generating figures of alignments using Microsoft Word and Microsoft PowerPoint.

Conclusions

In conclusion, we present several bioinformatics-based exercises that utilize amino acid and genomic DNA data from *Hoxa2* of zebrafish, human, and several other primate species and model organisms. SOPs are utilized in all exercises in training students. These exercises and their associated SOPs can be used for upper level undergraduate or graduate biology courses over the course of a semester. Since they are bioinformatics-based, they can be performed by students both within and outside laboratory environments. Further, the location of conserved CREs through comparative genomic analyses will allow students to generate hypotheses on the actual gene expression patterns of *Hoxa2* in zebrafish embryos. These exercises can aid in students' understanding of several complex concepts, including evolutionary constraint and purifying selection on genomic sequences, how mutations to CREs can generate morphological novelty in evolutionarily divergent species, and how mutations to CREs or DNA that gives rise to functional protein domains can lead to pleiotropic affects in several anatomical structures.

Table 6. Suggested sequence of steps to be performed for Exercise #3.

Step 1	Instructor generates antisense and sense *Hoxa2* riboprobes and tests these on zebrafish embryos prior to the beginning of the course. Manufactured antisense and sense riboprobes can be used by students for the WISH assay. Otherwise, students are trained on a SOP involved in producing sense and antisense *Hoxa2* riboprobes.
Step 2	Performance of WISH analysis (see BIO-006 SOP in the supplemental file).
Step 3	Mounting of embryos on microscope slides and digital photography of embryos.
Step 4	Modification of WISH images by instructor using Photoshop and distribution of images to students.
Step 5	Identification and labeling of rhombomeres and PAs by students using Microsoft PowerPoint. Students turn in assigned work for grading.

Acknowledgements

We would like to thank Dr. Pierre Le Pabic of University of North Carolina, Wilmington for providing us with fixed zebrafish embryos. Dr. Adam Davis would also like to thank Karen N. Donhauser, Director of Quality at Absorption Systems in Exton, PA, for her trust in Dr. Davis in promoting him to a Documentation Specialist when they worked together at Wyeth Pharmaceuticals in West Chester, PA. It was in this position that Dr. Davis learned how to write SOPs and train personnel.

Literature Cited

Amin, S., I.J. Donaldson, D.A. Zannino, J. Hensman, M. Rattray, M. Losa, F. Spitz, F. Ladam, C. Sagestrom, and N. Bobola. 2015. *Hoxa2* selectively enhances Meis binding to change a branchial arch ground state. Developmental Cell 32(3):265–277.

Baltzinger, M., M. Ori, M. Pasqualetti, I. Nardi, and F.M. Rijli. 2005. *Hoxa2* knockdown in *Xenopus* results in hyoid to mandibular homeosis. Developmental Dynamics 234(4):858–867.

Bhattacharya, J. 2015. Guidance for preparing standard operating procedures (SOPs). IOSR Journal of Pharmacy 5(1):29–36.

Benton, M.J. and P.C.J. Donoghue. 2007 Paleontological evidence to date the tree of life. Molecular Biology and Evolution 24(1):26–53.

Broughton, R.E., R. Betancur-R, C. Li, G. Arratia, and G. Orti. 2013. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLOS Currents 16:5.

Brudno, M., C.B. Do, G.M. Cooper, M.F. Kim, E. Davydov, E.D. Green, A. Sidow, and S. Batzoglou, 2003. LAGAN and Multi-LAGAN: Efficient Tools for Large-Scale Multiple Alignment of Genomic DNA. Genome Research 13(4):721–731.

Carroll, P.M., B. Dougherty, P. Ross-Macdonald, K. Browman, and K. FitzGerald. 2003. Model systems in drug discovery: Chemical genetics meets genomics. Pharmacology and Therapeutics 99(2):183–220.

Cattley, S. and J.W. Arthur. 2007. BioManager: The use of a bioinformatics we application as a teaching tool in undergraduate bioinformatics training. Briefings in Bioinformatics 8(6):457–465.

Chang, C.P., L. Brocchieri, W.F. Shen, C. Largman, and M.L. Cleary. 1996. Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. Molecular and Cellular Biology 167(4):1734–1745.

Cohen, J. 2003. Guidelines for establishing undergraduate bioinformatics courses. Journal of Science Education and Technology 12(4):449–456.

D'Costa, A. and I.T. Shepherd. 2009. Zebrafish development and genetics: Introducing undergraduates to developmental biology and genetics in a large introductory laboratory classroom. Zebrafish 6(2):169–177.

Davenne, M., M.K. Maconochie, R. Neun, A. Pattyn, P. Chambon, R. Krumlauf, and F.M. Rijli. 1999. *Hoxa2* and *Hoxb2* control dorsoventral patterns of neuronal development in the rostral hindbrain. Neuron 22(4):677–691.

Davidson, E.H. 2006. The Regulatory Genome: Gene Regulatory Networks in Development and Evolution. Amsterdam, The Netherlands Academic Press. 289 pp.

Davis, A., J.L. Scemama, and E.J. Stellwag. 2008. Japanese medaka *Hox* paralog group 2: Insights into the evolution of *Hox* PG2 gene composition and expression in the Osteichthyes. Journal of Experimental Zoology (Molecular and Developmental Evolution) 310(8):623–641.

Davis, A., M.C. Reubens, and E.J. Stellwag. 2016. Function and comparative genomics of *Hoxa2* gene *cis*-regulatory elements: evidence for evolutionary modification of ancestral core element activity. Journal of Developmental Biology 4(2):15.

Ditty, J.L., C.A. Kvaal, B. Goodner, S.K. Freyermuth, C. Bailey, R.A. Britton, S.G. Gordon, S. Heinhorst, K. Reed, Z. Xu, E.R. Sanders-Lorenz, S. Axen, E. Kim, M. Johns, K. Scott, and C.A. Kerfeld. 2010. Incorporating genomics and bioinformatics across the life sciences curriculum. PLOS Biology 8(8): e1000448. doi:10.1371/journal.pbio.1000448.

Floriano, W.B. 2008. A portable bioinformatics course for upper-division undergraduate curriculum in sciences. The International Union of Biochemistry and Molecular Biology 36(5):325–335.

Frasch, M., X. Chen, and T. Lufkin. 1995. Evolutionary-conserved enhancers direct region-specific expression of the murine *Hoxa-1* and *Hoxa-2* loci in both mice and Drosophila. Development 121(4):957-974.

Frazer, K.A., L. Pachter, A. Poliakov, E.M. Rubin, and I. Dubchak. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Research. 32(Web Server issue):W273–9

Gavalas, A., M. Davenne, A. Lumsden, P. Chambon, and F.M. Rijli. 1997. Role of *Hoxa-2* in axon pathfinding and rostral hindbrain patterning. Development. 124(19):3693–3702.

Gendron-Maguire, M., M. Mallo, M. Zhang, T. Gridley. 1993. *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. Cell 75(7):1317–1331.

Gilbert, S.F. and M.J.F. Barresi. 2016. Amphibians and fish. Pp. 333–378, *In* Developmental Biology. Eleventh Edition. S.F. Gilbert and M.J.F. Barresi, Eds. Sinauer Associates. Sunderland, MA. 912 pp.

Grammatopoulos, G.A., E. Bell, L., Toole, A. Lumsden, and A.S. Tucker. 2000. Homeotic transformation of branchial arch identity after *Hoxa2* overexpression. Development 127(24):2355–2365.

Hunter, M. and V.E. Prince. 2002. Zebrafish Hox paralogue group 2 genes function redundantly as selector genes to pattern the second pharyngeal arch. Developmental Biology 247(2):367–389.

Kimmel, C.B., W.W. Ballard, S.R. Kimmel, B. Ullman, and T.F. Schilling. 1995. Stages of embryonic development of the zebrafish. Developmental Dynamics 203(3):253–310.

Lampe, X., O.A. Samad, A. Guiguen, C. Matis, S. Remacle, J.J. Picard, F.M. Rijli, and R. Rezsohazy. 2008. An ultraconserved Hox-Pbx responsive element resides in the coding sequence of *Hoxa2* and is active in rhombomere 4. Nucleic Acids Research 36(10):3214–3225.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948.

LaRonde-LeBlanc, N.A. and C. Wolberger. 2003. Structure of Hoxa9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes and Development 17(16):2060–2072.

Le Pabic, P., J.L. Scemama, S.N. Brothers, and E.J. Stellwag. 2007. Comparative analysis of Hox paralog group 2 gene expression during Nile tilapia (*Oreochromis niloticus*) embryonic development. Development Genes and Evolution 217(11–12):749–758.

Le Pabic, P., J.L. Scemama, and E.J. Stellwag. 2010. Role of *Hox* PG2 genes in Nile tilapia pharyngeal arch specification: Implications for gnathostome pharyngeal arch evolution. Evolution and Development 12(1):45–60.

Maconochie, M.K., R. Krishnamurthy, S. Nonchev., P. Meier, M. Manzanares, P.J. Mitchell, and R. Krumlauf. 1999. Regulation of *Hoxa2* in cranial neural crest cells involves members of the *AP-2* family. Development 126(7):1483–1494.

Maconochie, M.K., S. Nonchev, M. Manzanares, H. Marshall, and R. Krumlauf. 2001. Differences in Krox20-dependent regulation of *Hoxa2* and *Hoxb2* during hindbrain development. Developmental Biology 233(2):468–481.

Maloney, M., J. Parker, M. LeBlanc, C.T. Woodard, M. Glackin, and M. Hanrahan. 2010. Bioinformatics and the undergraduate curriculum. CBE Life Sciences Education 9(3):172–174.

Mayor, C.., M. Brudno, J.R. Schwartz, A. Poliakov, E.M. Rubin, K.A. Frazer, L.S. Pachter, and I. Dubchak. 2000. VISTA: Visualizing Global DNA Sequence Alignments of Arbitrary Length. Bioinformatics 16(11):1046–1047.

McEllin, J.A., T.B. Alexander, S. Tümpel, L.M. Wiedemann, and R. Krumlauf. 2016. Analyses of fugu *Hoxa2* genes provide evidence for subfunctionalization of neural crest cell and rhombomere *cis*-regulatory modules during vertebrate evolution. Developmental Biology 409(2):530–542.

Nonchev, S., C. Vesque, M. Maconochie, T. Seitanidou, L. Ariza-McNaughton, M. Frain, H. Marshall, M.H. Sham, R. Krumlauf, and P. Charnay. 1996. Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20. Development 122(2):543–554.

Parker, H.A., M.E. Bronner, and R. Krumlauf. 2014. A *Hox* regulatory network of hindbrain segmentation is conserved to the base of vertebrates. Nature 514(7523):490–493.

Pasqualetti, M., M. Ori, I. Nardi, and F.M. Rijli. 2000. Ectopic *Hoxa2* induction after neural crest migration results in homeosis of jaw elements in *Xenopus*. Development 127(24):5367–5378.

Piper, D.E., A.H. Batchelor, C.P. Chang, M.L. Cleary, and C. Wolberger. 1999. Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. Cell 96(4):587–597.

Prince, V.E. and A. Lumsden. 1994. *Hoxa-2* expression in normal and transposed rhombomeres: independent regulation in the neural tube and neural crest. Development 120(4):911–923.

Prince, V. E., L. Joly, M. Ekker, and R.K. Ho. 1998. Zebrafish *Hox* genes: Genomic organization and modified colinear expression patterns in the trunk. Development 125(3)**:**407–420.

Rijli, F.M., M. Mark, S. Lakkaraju, A. Dierich, P. Dolle, and P. Chambon. 1993. A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. Cell 75(7):1333–1349.

Sarmah, S., G.W. Chismlll, M.A. Vaghan, P. Muralidharan, J.A. Marrs, and K.A. Marrs. 2016. Using zebrafish to implement a course-based undergraduate research experience to study teratogenesis in two biology laboratory courses. Zebrafish 13(4):293–304.

Scemama, J.L., Vernon, J.L., and Stellwag, E.J. 2006. Differential expression of *Hoxa2a* and *Hoxa2b* genes during striped bass embryonic development. Gene Expression Patterns 6(8):843–848.

Schmoldt, A., J. Forecki, D.R. Hammond, and A.J. Udvadia. 2009. Exploring differential gene expression in zebrafish to teach basic molecular biology skills. Zebrafish 6(2):187–199.

Sigrist, C.J.A., L. Cerutti, N. Hulo, A. Gattiker, L. Falquet, M. Pagni, A. Bairoch, and P. Bucher. 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Briefings in Bioinformatics 3:265–274.

Steiper, M.E. and N.M. Young. 2006. Primate molecular divergence dates. Molecular Phylogenetics and Evolution 41(2):384–394.

Tenorio, G. 2014. Using biological databases to teach evolution and biochemistry. Science in School 29(Summer 2014):30–34

Tümpel, S., M. Maconochie, L.M. Wiedemann, and R. Krumlauf, 2002. Conservation and diversity in the cis-regulatory networks that integrate information controlling expression of *Hoxa2* in hindbrain and cranial neural crest cells in vertebrates. Developmental Biology 246(1):45–56.

Tümpel, S., F. Cambronero, L.M. Wiedemann, and R. Krumlauf. 2006. Evolution of cis elements in the differential expression of two *Hoxa2* coparalogous genes in pufferfish (*Takifugu rubripes*). Proceedings of the National Academy of Sciences 103(14):5419–5424.

Tümpel, S., F. Cambronero, E. Ferretti, F. Blasi, L.M. Wiedemann, and R. Krumlauf, 2007. Expression of *Hoxa2* in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1. Developmental Biology 302(2):646–660.

Tümpel, S., F. Cambronero, C. Sims, R. Krumlauf, and L.M. Wiedemann. 2008. A regulatory module embedded in the coding region of *Hoxa2* controls expression in rhombomere 2. Proceedings of the National Academy of Sciences 105(51):20077–20082.

Tümpel, S., L.M. Wiedemann, and R. Krumlauf. 2009. Hox genes and segmentation of the vertebrate hindbrain. Current Topics in Developmental Biology 88:103–137.

Wash, P.D. and G.G. Freeman. 2013. BYOD – Engaging students using their own devices. National Social Science Technology Journal 3(1): http://nssa.us/tech_journal/volume_3-1/vol3-1_article8.htm.