Using Bioinformatics to Examine Bmb sequence conservation in fish.

by

Fatimat A. Olatunbosun

Submitted to Board of Study in Biology
School of Natural and Social Sciences
in partial fulfillment of the requirements
for the degree of Bachelor of Arts

Purchase College
State University of New York

May 2019

Sponsor: Dr. Elliott Abrams

Second Reader: Dr. Jonas

Table of Contents: Page number

Abstract :

Fertilization involves the fusion of the male and female gametes. In many vertebrates this process can be broken down into two main steps: (1) the fusion of the gamete plasma membranes followed by (2) the fusion of the corresponding pronuclei to ultimately form the one-celled zygote. Each of these steps are promoted by proteins embedded in the respective membranes. It has been recently shown that Bouncer (Bncr) protein is required for species specific fusion of the plasma membrane. Furthermore, the Brambleberry (Bmb) protein is required for the subsequent fusion of the pronuclei. The objective of this study is to identify conserved and non-conserved regions among the top 20 Bmb homologs in different fish species using a bioinformatics approach. We compare these results to the levels of conservation in Bncr, a known species specific fusogen. Bncr homologs are expected to diverge significantly, even among closely related fish species. As controls, we also include alignment data from both Ndc1 and Gp210, two components of the nuclear pore complex found in the nuclear envelope. Nuclear pore complex membrane proteins are not expected to diverge significantly among the top 20 homologs of each type. Collectively, these data will serve as a starting point into investigating the possible role of Bmb as a possible species-specific barrier in the process of fertilization.

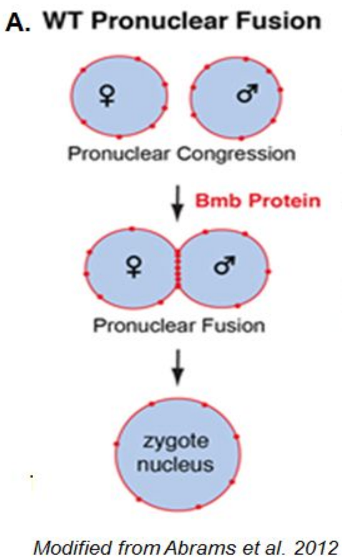
Introduction:

The ability to conceive children or young is known as fertility. Surprisingly, not much is known about fertility and the rate at which it may occur. There is a fertile window depending on the individual and this is the best time for conception. In general, the fertile window begins on the day of ovulation; which is the process when the luteinizing hormone (LH) triggers an ovary to release an egg during menstruation (americanpregnancy.org, 2019). Fertility varies based on genetics and this influences how many eggs a woman may be born with. Although hormones regulate the menstrual cycle many other factors can contribute to infertility (www.nichd.nih.gov/health, 2017). Some cases of infertility may be excessive alcohol and cigarette intake, having a low BMI, the shape of a woman's uterus and health problems such as diabetes and untreated diseases (mayoclinic.org, 2018). Even things like not incorporating exercise and a balanced healthy meal into your daily routine may affect a couples fertility.

Proteins are needed in order for fertilization to be successful in producing the outcome of a zygote. An oocyte-expressed species specific Ly6/uPAR protein otherwise known as Bouncer (Bncr), is a crucial fertilization factor in zebrafish. This protein functions in having a specific male gamete fertilize the female gamete, and if the male sperm is from a different species that the Bncr protein does not recognize then fertilization will not occur (fig1. b). It is not only required for sperm-egg interaction but it also allows cross-species fertilization between zebrafish and medaka; Mekada being a species that has diverged from zebrafish more than 200 million years ago. This divergence is an important factor in why Bncr serves as a key determinant for species specific fertilization. Without Bncr the sperm can not enter into the egg (Schoft, Vera K et al., 2003). It is because of this species-specific protein that we compare it to Bmb. We plan on determining where the Bmb protein falls on the spectrum level of conservation. We do this by comparing and contrasting Bmb to other proteins that may be similar and proteins that may be completely diverged from it. Bmb is a protein that is involved in nuclear envelope fusion. This is known based on studies that have been successfully experimented. Studies by Abrams et al have proven that there are specific proteins necessary for cell fusion that occur in the zygote stage of zebrafish. Using a cloned Bmb mutant gene, Bmb is identified in having a 27% similarity in fusion characteristics to that of kar5p, which is a protein that aids in the process of nuclear fusion in yeast or fungus (Abrams et al,. 2012).

Since we know that Bmb aggregates at the pronuclear interface, we also compare this to other proteins that function in the transporting of proteins and molecules on the nuclear envelope. Figure 1 a-d show the visualization of the 3 groups that will be compared on the levels of conservation. The aggregation at the interface plate of both the maternal and paternal gamete fuse with the help of Bmb. This fusion is crucial in pronuclear fusion. Figure 1b magnifies the maternal eggs cell membrane. With species specific Bncr, fertilization occurs at the plasma membrane. The specific sperm is identified by Bncr which then allows it to enter the membrane. If there is a foreign sperm that is not species-specific it will not fuse with the cell membrane. This is why the protein is named Bncr, since it has traits acting as a bodyguard keeping foreign proteins out of the plasma membrane. Figure 1e shows the mouse oocyte that is surrounded by the thick layer of the zona pellucida (zp), as well as the three proteins (ZP1,ZP2, and

ZP3) that aid in the prevention of polyspermy (Dun, Matt et al., 2010). In mice the zp also plays a crucial role in process of fusion. This is a specialized extracellular matrix surrounding the developing oocyte (egg, ovum) within each follicle in the ovary. This thick matrix is thought to be formed by secretions from the oocyte and the follicle granulosa cells. In humans, oocytes consist of four types of zona pellucida glycoproteins ZP1, ZP2, ZP3 and ZP4 which play different roles in fertilization (Mold, D et al., 2009).The thick membrane of the zona pellucida plays an important role in the process of sperm and egg recognition and its protein function allows species-specific fertilization, which can be compared to Bcnr. Its thick membrane forms a barrier around the egg and this barrier prevents polyspermy, and enables the acrosome reaction for the successful adhesion and penetration of the sperm cell (Bukovsky et al., 2010). Once the sperm penetrates the zona pellucida it initiates a process called the zona reaction. This solidifies and changes the chemical properties of the zona pellucida. The GALT is located on the sperm, and the ZP3 is also located on the zona pellucida. This interaction acts like Bncr with its species-specific behavior. Once this interaction occurs the sperm head secretes the acrosin which are enzymes that chew away at the zona pellucida in order for it to enter the membrane. After this, there is cortical granule exocytosis; which are vesicles that remain at the surface of the pre-fertilized egg and when the sperm comes in exocytosis reacts also preventing polyspermy. This is a double bodyguard acting method that further protects and stabilizes the control of only one sperm penetrating the egg. This chain reaction stops polyspermy from occurring and resulting to a polyploid egg, genetically destructing the organism. These are the first two layers of the zona pellucida. As for the third layer, the ADAMS interact with integrins and this is a transmembrane protein located on the plasma membrane. These interactions allow the transfer of the pronucleus into the cytoplasm, which also requires a protein called tetraspanin which needs to be on the egg. The cortical granules will release proteases and glycosidases. They disrupt any sperm-egg interactions that can cause polyploidy. When the sperm enters it will spike calcium and the release of calcium sends out the proteases and glycosidases into the extracellular environment and this chemical process also acts to block polyspermy (K. Gupta et al., 2012).



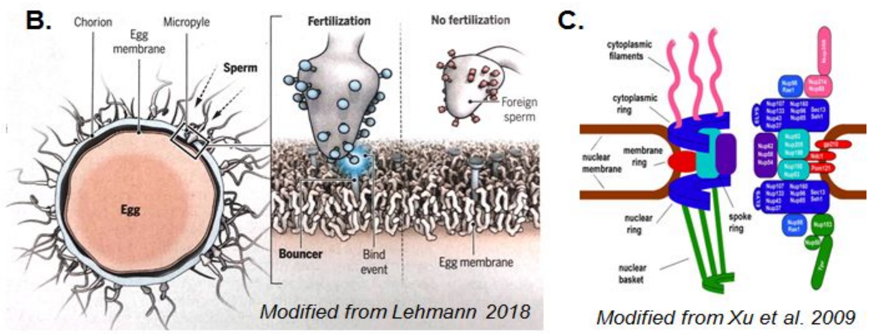


Figure1: Brambleberry and Bouncer localization and function during fertilization. A. Maternal Brambleberry (Bmb, red) is a transmembrane protein that is localized to nuclear cortex and is required for pronuclear fusion in zebrafish. **B.** Maternal Bouncer (Bcnr) is a species-specific membrane protein that is required for plasma membrane fusion during fertilization. **C.** Ndc1 and gp210 (highlighted in red) are transmembrane components of the nuclear pore complex (NPC). These NPC were included in the study as a control for evaluating regional conservation of proteins among the species examined.

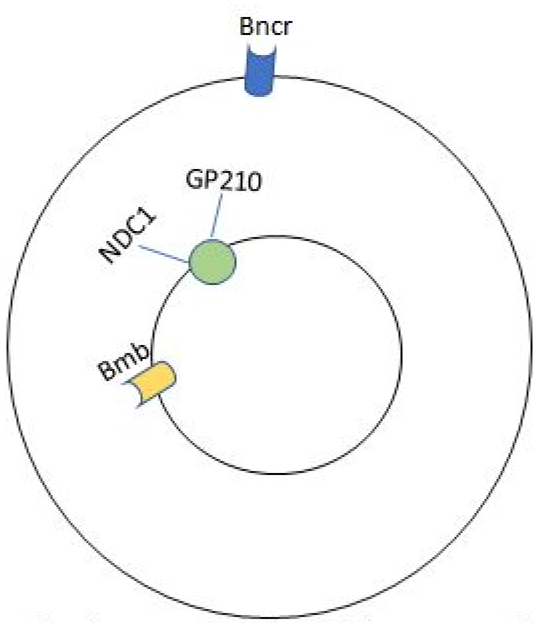


Figure 1d: Cell containing Bmb, npc's (Pom121, Ndc1, and Gp210) and species-specific bncr protein. Bmb and npc's allow for the fusing of the nuclear envelope membrane while bncr allows the fusing of the plasma membrane.

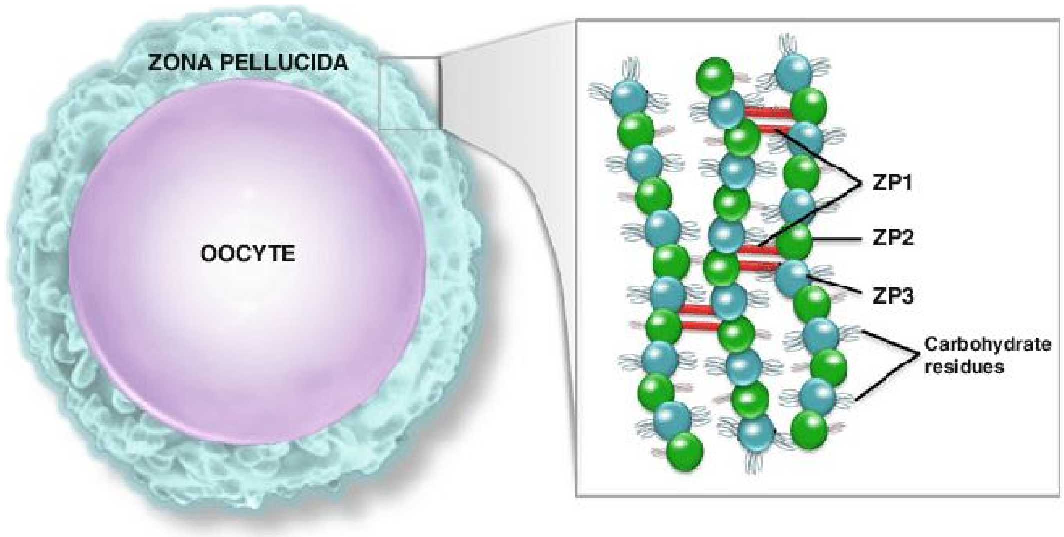


Figure 1e:Zona pellucida in mice; Image shows a representation of the mouse oocyte surrounded by the zona pellucida (zp). This zp functions in allowing the entrance of sperm into the cell membrane by species-specificity. This involves gamete interaction, and it also prevents polyspermy from occuring.

By separating the grouped proteins based on their function we can then compare them to Bmb. Another group of proteins are the nuclear pore complexes (npc's), which are transmembrane proteins and they function in the portal access of molecules in and out of the nucleus and they also anchor npc's. Nuclear pore complexes (NPC's) as seen in fig 1. C are a protein-lined channel in the nuclear envelope that regulate the transportation of molecules between the nucleus and cytoplasm. This occurs in eukaryotic cells and serves as a barrier between the cytoplasm and nucleus to prevent harm to genetic material housed in the nucleus. There are approximately 1,000 NPC's in the nuclear envelope. This varies from cell to cell and it also varies depending on the life cycle stages in organisms (Nature.com, 2014). Each NPC has about 456 individual protein molecules that are composed of 34 distinct proteins and each contain about 30 nucleoporins. Out of the wide variety of proteins, they were narrowed down in order to have a control group which is different in regards to similarity. Also these two npc's have high levels of conservation which will be discussed and and seen in figure 3. The two transmembrane proteins (tp) are as follows, Gp210, and Ndc1. GP210 is an essential trafficking regulator. It anchors the pore complex to the nuclear membrane and GFP protein tagging reveals its primarily located on the luminal side of double layer membrane at the pore. Ndc1 is a conserved tp. It plays a key role in new assembly and insertion of NPC in the nuclear envelope. It also forms a link between the nuclear envelope membrane and soluble nucleoporins, which anchors the NPC in the membrane (Uniprot.org, 2018).

In order to figure out where Bmb falls on this spectrum we have to use certain tools that will guide us to accept our hypothesis. We want to compare sequence conservation amongst three categories. We take the bioinformatics approach and we use an important model organism called *Danio rerio* (zebrafish), which is ideal for genetic and developmental studies. Our study examines the possibility of Bmb having a species-specific function. Mapping out its conservation levels, the bioinformatics approach is chosen in order to analyze the structure of Bmb's protein sequence while aligned to the top 20 similar fish species. Knowing that these fish will never be 100% similar, they are aligned in order to find the specific conserved and non-conserved regions. Mapping out the top 20 homologs allows us to further investigate whether Bmb serves as a protein that is species-specific or not. The idea is that if there are domains in low homology this is where the species specificity will be. If Bmb is species specific, then it can then be linked to Bncr; which functions in being species specific for the fusion of the plasma membrane. Two TM npc proteins were chosen to serve as controls, since they anchor non-transmembrane component to the complex. With more information using the bioinformatics approach we can research ways to link species specific proteins at the nuclear envelope level.

Methods:

This study employs a bioinformatics approach to analyze complex biological protein data. Bmb is aligned and compared to separate proteins in order to visualize the conserved and nonconserved areas in their sequences. The function of each protein involves either the fusing of the plasma membrane or nuclear envelope fusion. In this experiment species-specific protein Bncr functions in the fusing of the plasma membrane. While the control group of npc's serve as a gateway of materials to enter and exit the nucleus. There are many proteins that are specific to mammals, fish, frogs and other vertebrates but this research is limited to specific proteins that involve fertilization. In order to identify the top 20 homologs corresponding to the proteins we use NCBI BLAST, which is a bioinformatics tool. Using BLAST, each protein (Bncr, Bmb, Gp210, and Ndc1) are used to create paralogs. It is important to note that in order to have a successful paralog the genus of fish needs to be different in order to create a source of variety since we are looking to compare and contrast protein sequences. If the species is the same the bioinformatics website Praline will not be able to run the alignment because of interferences that disrupt species variety. As well as comparing the sequence databases BLAST also calculates the statistical significance of the protein or lack thereof (blast.ncbi.nlm.nih.gov).

PRofile ALIgNEment (PRALINE) is the second database used after BLAST in order to customize multiple sequence alignments. In addition to a number of available alignment strategies, PRALINE integrates information from database homology searches to generate a homology-extended multiple alignment. It then integrates the information from a separate database homology search which then generates a homology-extended multiple alignment creating a color coded conservation map. The red areas express the highly conserved area whereas the blue expresses the non-conserved areas. Everything in between the colors of red to blue vary from conserved to unconseved (Simossis, V. A., & Heringa, J., 2005). Before finding the top 20 homologs the first task was to search for the Bmb protein, since this is the protein that is used throughout the experiment to link us to finding a species specific area in the protein sequences. Using the Bmb protein sequence, it is then placed into the BLAST tool website and aligned against the four proteins of interest as stated earlier. Many species of fish are aligned and separated based on their sequence identity to Bmb. We only use the top 20 species of fish that are most similar. There were many protein sequences that were excluded during research while using certain tool websites such as BLAST. The proteins on this website are mostly predicted, because of this the data given is raw genomic DNA from which the program then predicts the sequence. Therefore there were certain proteins in that were missing certain sections in either the n or c-terminus of the amino acid (aa) sequence so this is why they were marked as our exclusions. When the proteins were run on Praline a blank space is what was seen and for this reason is why it is important to check and run many alignments making sure that the sequences are not unpredicted in order to get the full homologous sequence. When running a search on BLAST the results will give a variety of species in the order of their similarity to the protein that you are aligning. The similarity percentage helps to keep track of how similar the species protein is to the aligned protein.

Once the top twenty proteins were obtain, an alignment using Praline was performed to identify homologous vs. non homologous regions. This site creates a colorful warm-scale conservation map along with a number color log showing the conservation levels from conserved to unconserved. This shows the entire protein sequences taken from BLAST from the N-terminus to the C-terminus. Comparing the evolutionary conservation, altogether with the five proteins aligned there were 42 different species of fish represented. The alignment is separated by proteins bmb, bmb related, bncr, gp210, and ndc1. The colors represent the conservation of the specific region of the protein. This scale is numbered from 1-10. 1 is labeled non-conserved, while red is labeled as very conserved. The colors of green and yellow represent the intermediate values of conservation. In each of these five homology groups there are certain domains that are represented within the color of the conservation map. These domains represent regions that are highly conserved within fish species[MOU1] .
Next, I generated phylogenetic trees corresponding to highly conserved regions in the proteins. In order to create the schematics seen in fig 3. The praline maps are used in order to calculate and average the scaled numbers from the n to c-terminus. Each compiled bar represents one protein sequence. This was done by averaging the regions of similar colors on the scale of 1-10 which was then divided by 100 in order to get the exact color number for that specific region. By carefully separating those regions you can then see the drift (fig 3) in each compiled schematic that drifts from very conserved to un-conserved.

Tree diagrams, also known as phylogenetic trees have been used in evolutionary biology since the time of Charles Darwin. Phylogenetic trees are diagrams that depicts the lines of evolutionary descent of different species, organisms, or genes from a common ancestor. We also used these trees as seen in figure 4d to show the significant homologies in each of the specific species that are aligned the proteins. There are clades that cluster showing the similarity and the separation of the fish, this proves that even though the fish are similar to one another they are not 100% identical hence evolution drifting them apart. Even with similar identities in protein structure homologies the fish still break off in differences. The split in the different clades (fig.4) shows how there is a split that spans in the chromosome. One important note to remember is that the fish aligned do not have the same protein structure because if they did then the results would be a narrow spectrum and that data would not be broad enough to research for species-specificity, but if they are similar they will contain certain sequences which will ultimately result in conservation. This experiment is meant to test the boundaries of visualizing conserved and non-conserved regions based on a paralog protein alignment. Only then can you physically see if the structure of the proteins are similar which will then force one to draw the conclusion that Bmb is species-specific.

Results:

Since fertilization serves as an important factor for producing a one-celled zygote it is crucial to know what specific proteins aid in structure and fusing of both the plasma membrane as well as the nuclear envelope. The first step was to identify the top 20 homologs that correspond to the five proteins listed in table 1. In order to do this NCBI BLAST is used. This is the first step to the bioinformatics approach that was used in order to find similar fish to that of the Bmb protein. First we take the Bmb protein sequence and place it into the BLAST website. Our study focused on the nuclear envelope fusion protein, Bmb. This study also focuses on fish. To stray away from this specific specie would lead us to find results that would diverge us from our goal of linking B
mb as having regions that are species-specific. The main question of species specificity involving Bmb also allows us to find certain domains of the five proteins and align them as seen in table 1. The sequence identification numbers are also placed into the table, and when clicked it references you back to BLAST and shows the actual protein sequence of the specific fish. This table separates the proteins and places all species of fish homologies under one table. This also allows us to visualize and pinpoint the specific areas that are similar between the 42 species of fish.

Bmb	TM Domain	CC Domain	Sequence Id #
Danio rerio (zebrafish)	1 (363-386), 2 (415-432)	1 (250-380)	NP_001257483.1
Cyprinus carpio (European carp)	1 (368-385), 2 (410-431)	0	KTG45706.1
Sinocyclocheilus anshuiensis (cavefish)	1 (374-391), 2 (416-437)	0	XP_016350013.1
Seriola dumerili (greater amberjack)	1 (372-395), 2 (423-440)	0	XP_022608732.1
Sinocyclocheilus rhinocerous (ray-finned fish)	1 (368-385), 2 (414-431)	0	XP_016378109.1
Sinocyclocheilus grahami (golden line barbel cyprinid fish)	1 (374-391), 2 (420-437)	0	XP_016094067.1
Ictalurus punctatus (catfish)	1 (383-406), 2 (434-452)	0	XP_017341701.1
Pygocentrus nattereri (red-bellied piranha)	1 (372-395), 2 (420-441)	0	XP_017554919.1
Astatotilapia calliptera (eastern happy)	1 (365-385), 2 (416-433)	0	XP_026004148.1
Monopterus albus (white ricefield eel)	1 (367-390), 2 (418-435)	0	XP_020457605.1
Mastacembelus armatus (tire track eel)	1 (381-401), 2 (423-443)	0	XP_026179037.1
Salmo salar (atlantic salmon)	1 (324-449), 2 (480-498)	0	XP_013978724.1
Salvelinus alpinus (arctic char)	1 (416-441), 2 (472-490)	0	XP_023859660.1
Labrus bergylta (Ballan wrasse)	1 (368-391), 2 (419-436)	0	XP_020492939.1
Pangasianodon hypophthalmus (iridescent catfish)	1 (371-394), 2 (423-440)	0	XP_026774666.1
Electrophorus electricus (electric eel)	1 (370-393), 2 (421-439)	0	XP_026856837.1
Oreochromis niloticus (nile tilapia)	1 (371-394), 2 (422-439)	0	XP_005458344.1
Perca flavescens (yellow perch)	1 (368-391), 2 (426-446)	0	XP_028442731.1
Anabas testudineus (the climbing perch)	1 (375-394), 2 (418-435)	0	XP_026201587.1
Xiphophorus maculatus (southernh playtfish)	1 (366-389), 2 (417-435)	0	XP_023196050.1
Oryzias melastigma (ricefish)	1 (365-388), 2 (416-433)	0	XP_024134239.1

Bmb Related	TM Domain	CC Domain	Sequence Id #
Danio rerio (zebrafish)	1 (228-247), 2(250-271), 3 (295-315)	1 (320-360)	XP_693095.4
Carassius auratus (goldfish)	1(230-249), 2(252-274), 3(297-315)	1 (320-360)	XP_026124937.1
Sinocyclocheilus grahami (golden line barbel cyprinid fish)	1(229-249), 2(253-272), 3(297-316), 4(376-395)	1 (320-360)	XP_016089552.1
Cyprinus carpio (European carp)	1 (235-253), 2 (259-286), 3 (302-322)	1 (320-360)	XP_018085225.1
Labeo rohita (South Asian rohu)	1 (236-254), 2 (259-284), 3 (302-323)	1 (320-360)	XP_020648687.1
Astyanax mexicanus (Mexican tetra)	1(229-248), 2(251-273), 3(296-316)	1 (320-360)	KTG46871.1
Pygocentrus nattereri red-bellied piranha)	1(230-249), 2(252-274), 3(297-317)	1 (320-360)	RXN10293.1
Ictalurus punctatus (catfish)	1(231-248), 2(255-275), 3(298-316)	1 (320-360)	XP_007245898.2
Oryzias latipes (Japanese rice fish)	1(233-250), 2(257-277), 3(300-320)	1 (320-360)	XP_017574954.1
Oreochromis niloticus (nile tilapia)	1(249-268), 2(273-293), 3(316-336)	1 (330-370)	XP_017324046.1
Tachysurus fulvidraco (yellowhead catfish)	1(237-254), 2(257-277), 3(304-322)	1 (270-320)	XP_004075992.1
Salmo salar (Atlantic salmon)	1(243-260), 2(310-328)	1 (340-370)	XP_025757437.1
Clupea harengus (Atlantic herring)	1(231-250), 2(255-275), 3(295-316)	1 (320-360)	XP_027006963.1
Oncorhynchus kisutch (coho salmon)	1(235-252), 2(257-277), 3(301-322)	1 (320-360)	XP_026532061.1
Astatotilapia calliptera (Eastern happy river beam)	1(233-250), 2(257-277), 3(300-318)	0	XP_023691658.1
Salvelinus alpinus (Arctic char)	1(248-265), 2(271-291), 3(316-334)	1 (330-370)	XP_014002992.1
Esox lucius (Northern Pike)	1(233-250), 2(256-277), 3(298-318)	0	XP_012681540.1
Poecilia formosa (amazon molly)	1(245-262), 2(312-330)	1 (340-370)	XP_020329547.1
Xiphophorus couchianus (Monterrey platyfish)	1(233-250), 2(256-277), 3(298-318)	0	XP_026009447.1
Acanthochromis polyacanthus (spiny chromis)	1(233-253), 2(252-272), 3(295-315)	0	XP_023993959.1

Bncr	TM Domain	CC Domain	Sequence Id #
Danio rerio (zebrafish)	1 (79-97), 2 (109-125)	0	NP_001352655.1
Sinocyclocheilus grahami (golden line barbel cyprinid fish)	1(120-146)	0	1. XP_016146958.1
Cyprinus carpio (European carp)	1(119-137)	0	XP_018955736.1
Cynoglossus semilaevis (Bengal tongue)	1(109-129)	0	XP_008310643.1
Paralichthys olivaceus (olive flounder)	1(106-125)	0	XP_019934301.1
Scophthalmus maximus (turbot)	1(111-128)	0	1. AWP08232.1
Labrus bergylta (Ballan wrasse)	1(111-128)	0	1. XP_020514534.1
Larimichthys crocea (yellow croaker)	1(115-132)	0	XP_027128866.1
Boleophthalmus pectinirostris (mudhopper)	1(107-131)	0	1. XP_020790982.1
Hippocampus comes (tiger tail seahorse)	1(105-128)	0	1. XP_019712504.1
Paramormyrops kingsleyae	1(118-136)	0	XP_023681726.1
Oryzias javanicus (japanese ricefish)	1(107-124)	0	1. RVE71661.1
Esox lucius (Northern Pike)	1(110-129)	0	1. XP_010900260.1
Takifugu rubripes (japanese pufferfish)	1(110-131)	0	XP_011605859.1
Scleropages formosus (asian roho)	1(107-125)	0	XP_018602791.1
Lepisosteus oculatus (spotted gar)	0	0	XP_015207687.1
Oryzias melastigma (ricefish)	0	0	XP_024137197.1
Salmo salar (Atlantic salmon)	1(111-130)	0	XP_013981439.1
Poecilia formosa (amazon molly)	1(22-43), 2(110-126)	0	XP_005878225.1
Perca flavescens (yellow perch)	1(54-74), 2(107-123)	0	1. OCA29924.1
Kryptolebias marmoratus (mangrove killifish)	0	0	1. XP_017261251.1

NDC1	TM Domain	CC Domain	Sequence Id #
Danio rerio (zebrafish)	1 (61-80), 2 (113-130), 3 (158-174), 4 (218-235), 5 (263-281), 6 (580-599)	1 (360-385)	NP_956237.1
Cyprinus carpio (European carp)	1 (64-83), 2 (116-132), 3 (161-177), 4 (214-232), 5 (266-284), 6 (583-602)	0	KTF89731.1
Sinocyclocheilus anshuiensis (cavefish)	1 (61-80), 2 (113-130), 3 (158-174), 4 (208-229), 5 (263-281), 6 (580-599)	0	XP_016323463.1
Carassius auratus (goldfish)	1 (64-83), 2 (116-133), 3 (161-177), 4 (214-232), 5 (266-284), 6 (583-602)	0	XP_026141838.1
Sinocyclocheilus grahami (golden line barbel cyprinid fish)	1 (277-294), 2 (322-338), 3 (390-408), 4 (442-460), 5 (742-761)	1 (10-90), 2 (90-160), 3 (200-220)	ROL44648.1
Astyanax mexicanus (Mexican tetra)	1 (60-81), 2 (116-132), 3 (159-175), 4 (215-234), 5 (264-282)	0	XP_022525720.1
Electrophorus electricus (electric eel)	1 (65-83), 2 (116-132), 3 (159-175), 4 (264-282), 5 (581-600)	0	XP_026864394.1
Pygocentrus nattereri (red-belilied pirahna)	1 (56-81), 2 (116-132), 3 (159-175), 4 (264-282), 5 (586-605)	0	XP_017561805.1
Larimichthys crocea (yellow croaker)	1 (70-89), 2 (119-137), 3 (164-182), 4 (219-242), 5 (269-289), 6 (587-606)	0	XP_027144646.1
Ictalurus punctatus (channel catfish)	1 (65-83), 2 (113-131), 3 (156-177), 4 (257-281), 5 (416-435), 6 (575-594)	0	XP_017346622.1
Lates calcarifer (barramundi)	1 (64-80), 2 (113-131), 3 (155-173), 4 (210-233), 5 (262-280), 6 (579-598)	0	XP_018533462.1
Boleophthalmus pectinirostris (mudhopper)	1 (58-83), 2 (113-131), 3 (157-175), 4 (212-230), 5 (265-285), 6 (579-598)	0	XP_020784831.1
Tachysurus fulvidraco (yellowhead catfish)	1 (65-83), 2 (113-131), 3 (158-177), 4 (257-281)	0	XP_027012489.1
Clupea harengus (Atlantic herring)	1 (65-83), 2 (113-131), 3 (157-178), 4 (264-282)	0	XP_012692487.1
Monopterus albus (asian swamp eel)	1 (95-111), 2 (146-162), 3 (187-207), 4 (223-244), 5 (240-263), 6 (292-308), 7 (606-627)	0	XP_020447325.1
Seriola lalandi dorsalis (California yellowtail)	1 (65-87), 2 (120-141), 3 (162-180), 4 (217-240), 5 (269-287), 6 (585-604)	0	XP_023263189.1
Oreochromis niloticus (Nile tilapia)	1 (61-80), 2 (113-132), 3 (155-173), 4 (210-233), 5 (262-280), 6 (575-594)	0	XP_005476678.1
Seriola dumerili (greater amberjack)	1 (65-87), 2 (120-141), 3 (162-180), 4 (217-240), 5 (269-287), 6 (585-604)	0	XP_022613646.1
Labrus bergylta (Ballan wrasse)	1 (64-80), 2 (86-104), 3 (116-132), 4 (154-173), 5 (259-277), 6 (575-594)	0	XP_020515984.1
Stegastes partitus (Bicolor damselfish)	1 (62-80), 2 (113-132), 3 (155-173), 4 (210-233), 5 (262-280), 6 (578-597)	0	XP_008275392.1
Anabas testudineus (the climbing perch)	1 (64-83), 2 (113-131), 3 (155-173), 4 (210-233), 5 (262-280), 6 (577-596)	0	XP_026231184.1

GP210	TM Domain	CC Domain	Sequence Id #
Danio rerio (zebrafish)	1 (1,016), 2 (1,093), 3 (1,456), 4 (1,739-1,761), 5 (1,783-1,800)	0	XP_002667606.3
Sinocyclocheilus anshuiensis (cavefish)	1 (1,016-1,032), 2 (1,093-1,119) 3 (1,456-1,480), 4 (1,739-1,761), 5 (1,783-1,800)	0	XP_016336749.1
Cyprinus carpio (European carp)	1 (1,175-1,193), 2 (1,382-1,398), 3 (1,744-1,766), 4 (1,798-11,816)	0	XP_018921767.1
Carassius auratus (goldfish)	1 (941-964), 2 (1,380-1,396), 3 (1,611-1,631), 4 (1,796-1,814)	0	XP_026097199.1
Labeo rohita (South Asian rohu)	1 (147-165), 2 (933-956), 3 (1,767-1,785)	0	RXN24121.1
Pygocentrus nattereri (red-belilied pirahna)	1 (841-861), 2 (915-933), 3 (948-971), 4 (1,123-1,141), 5 (1,185-1,208), 6 (1,755-1,766), 7 (1,804-1,822)	0	XP_017553748.1
Astyanax mexicanus (Mexican tetra)	1 (845-870), 2 (916-934), 3 (949-972), 4 (1,124-1,142), 5 (1,186-1,209), 6 (1,756-1,774), 7 (1,806-1,824)	0	XP_022532509.1
Electrophorus electricus (electric eel)	1 (844-865), 2 (920-936), 3 (951-974), 4 (1,118-1,144), 5 (1,188-1,212), 6 (1,441-1,458), 7 (1,482-1,505), 8 (1,807-1,825)	0	XP_026857578.1
Ictalurus punctatus (channel catfish)	1 (952-975), 2 (1,127-1,145), 3 (1,808-1,826)	0	XP_017343380.1
Clupea harengus (Atlantic herring)	1 (493-518), 2 (871-893), 3 (1,808-1,827)	0	XP_012682893.1
Tachysurus fulvidraco (yellowhead catfish)	1 (641-663), 2 (951-975), 3 (1,126-1,144), 4 (1,802-1,820)	0	XP_027015266.1
Salvelinus alpinus (Arctic char)	1 (480-500), 2 (493-517), 3 (849-869), 4 (1,131-1,149), 5 (1,397-1,417), 6 (1,765-1,784), 1,812-1,832)	0	XP_023846423.1
Salmo salar (Atlantic salmon)	1 (493-513), 2 (506-530), 3 (862-882), 4 (938-954), (969-992), 6 (1,772-1,792), 7 (1,825-1,844)	0	XP_014000915.1
Oncorhynchus mykiss (rainbow trout)	1 (486-506), 2 (499-523), 3 (962-985), 4 (1,137-1,155), 5 (1,403-1,423), 6 (1,771-1,790), 7 (1,818-1,838)	0	XP_021471220.1
Seriola dumerili (greater amberjack)	1 (496-520), 2 (849-869), 3 (1,131-1,149), 4 (1,342-1,365), 5 (1,765-1,784), 6 (1,814-1,833)	0	XP_022594847.1
Esox lucius (Northern Pike)	1 (486-506), 2 (499-523), 3 (960-983), 4 (1,193-1,223), 5 (1,401-1,421), 6 (1,752-1,772), 7 (1,816-1,836)	0	XP_010893608.1
Lates calcarifer (Barramundi)	1 (500-525), 2 (853-873), 3 (1,135-1,153), 4 (1,346-1,369), 5 (1,769-1,788), 6 (1,816-1,837)	0	XP_018556619.1
Paramormyrops kingsleyae	1 (929-948), 2 (963-986), 3 (1,053-1,072), 4 (1,823-1,841)	0	XP_023660832.1
Mastacembelus armatus (tire track eel)	1 (497-522), 2 (850-870), 3 (1,132-1,150), 4 (1,343-1,366), 5 (1,764-1,785), 6 (1,815-1,834)	0	XP_026188065.1
Anabas testudineus (the climbing perch)	1 (499-524), 2 (852-873), 3 (1,134-1,152), 4 (1,196-1,220), 5 (1,345-1,363), 6 (1,402-1,420), 7 (1,768-1,787), 8 (1,815-1,834)	0	XP_026223609.1
Scophthalmus maximus (turbot)	1 (300-324), 2 (493-519), 3 (953-976), 4 (1,128-1,146), 5 (1,339-1,362), 6 (1,758-1,781), 7 (1,809-1,830)	0	AWP09921.1

Table 1. Protein homology domain alignment used in this study. Corresponding species with common names are indicated. Transmembrane (TM) and coil-coil (CC) domains are specified at the specified amino acid positions (in parentheses).

The domains that are listed are as follows; (refer to table 1) transmembrane domains (tm domain), and coil-coil domains (cc domain). These two domains were found on the two sites (transmembrane domain server, coil coil server). These sites allowed us to see how similar each specific species of fish really were to one another, or if they were similar at all. A transmembrane protein (TP) is a type of integral membrane protein that spans the entirety of the cell membrane to which it is permanently attached. Transmembrane domains (TMDs) are a set of nonpolar amino acid residues. TMDs usually consist of α helices. The peptide bond is intrinsically polar and can form internal hydrogen bonds between carbonyl oxygens and amide nitrogens, or either of these may be hydrated. Within the lipid bilayer, where water is essentially excluded, peptides usually adopt the α helical configuration that maximizes their internal hydrogen bonding. A length of helix of 18–21 amino acid residues is sufficient to span the usual width of a lipid bilayer.

As for coiled-coil is a structural motif in proteins in which 2–7 alpha-helices are coiled together like the strands of a rope (dimers and trimers are the most common types). Many coiled coil-type proteins are involved in important biological functions such as the regulation of gene expression. The coil-coil domains are important in protein-protein interactions (Mason et. al., 2004). Bmb may interact with this domain or it may even interact with itself, or the Bmb related protein since it is mostly similar. Starting from the two similar proteins that were aligned (Bmb and Bmb related) in table 1, there are areas in the protein sequence that show high similarity. The zebrafish in the Bmb protein has two tm domains that range from (363-386) and (415-432). The zebrafish in the Bmb related protein has three tm domains which range from (228-247), (250-271), and (295-315). If you look at just these two differences, you can note that Bmb only has two transmembranes while Bmb related has three. This is the beginning of another hypothesis that can be drawn out that both Bmb and Bmb related are quite different from one another although they are conserved and similar to one another. Another important point gathered by the table allows us to come to the conclusion that in each of the 20 species of fish in the Bmb protein the ranges of the transmembrane domains are not far apart from one another, but they also are in fact not similar to one another. When putting together this data it was a surprise to see that there is not a single transmembrane domain with an identical mate. This means that even though these fish were labeled as being the top 20 aligned against that of Bmb, this doesn't necessarily mean that they will have the same domains.

The findings are quite the opposite when you look at how different the transmembrane domains are. With a few exceptions there are fish that have similar transmembrane domains but they aren't the same exact number. Examples would be the European carp and the ray finned fish. Although both of their first transmembrane domains are the same (363-386), their second transmembrane domains differ (410-431) and (414-431). In Bmb related the transmembrane domains in each of the fish vary drastically. Some of the species of fish have four domains, while most have three domains. Moving down towards the less identical fish, which are the fish that are located at the bottom of the Bmb table, the Atlantic salmon and the amazon Molly both have two transmembrane domains. The variation of domains shows that these fish are not 100% identical to Bmb, and this makes sense because the fish are not the same when compared to that of Bmb. 42 different species of fish were chosen for this study. The fish are meant to be different in order to get a broad spectrum of species to compare zebrafishes Bmb protein to. The species-specific Bncr protein has either one or two transmembrane proteins. Differences in this un-conserved protein lie in the Asian rohu the spotted gar and the mangrove killifish with a tm domain of zero. There are two fish that are 100% identical when comparing their tm domains which are the turbot and the ballan wrasse which have a range of (11-128). Since both of these fish are in order then conclusion can be drawn that these two fish are very similar to one another hence them being right after one another. As for the control groups which are the npc's, the tm domains for these two proteins vary drastically. With Ndc1 containing mostly 6 domains each their values range from (61-630). While Gp210 has mostly 8 tm domains. The range in tm domains differ drastically yet none of the domains from the species of fish are the same in either two npc's.

The cc domains can also be viewed in the third column of table 1. In the first cc row labeled for zebrafish this column surprisingly only has the protein as having one cc domain with a range of (250-380), while the other 20 species of fish do not have any cc domains in their protein sequence. When compared to the second region of the table (Bmb related), the species of fish under this protein have one cc domain with a range of (320-360) with the acception of the Eastern happy river beam, the northern pike, Monterrey platyfish, and the spiny chromis having no cc domain. The yellow head catfish has the lowest domain with a range of (340-370). Species-specific Bncr does not have a single cc domain in all 20 species of fish, and this is the same result for the npc Gp210. Last but not least Ndc1 has 2 cc domains represented in two fish. The first is zebrafish and the range is from (360-385) while the golden line barbel Cyprinid fish is represented as having 3 cc domains ranging from (10-90), (90-160), and (200-220). Now, the domains found on each protein are then taken from the table and mapped out onto a cartoon compiled schematic.

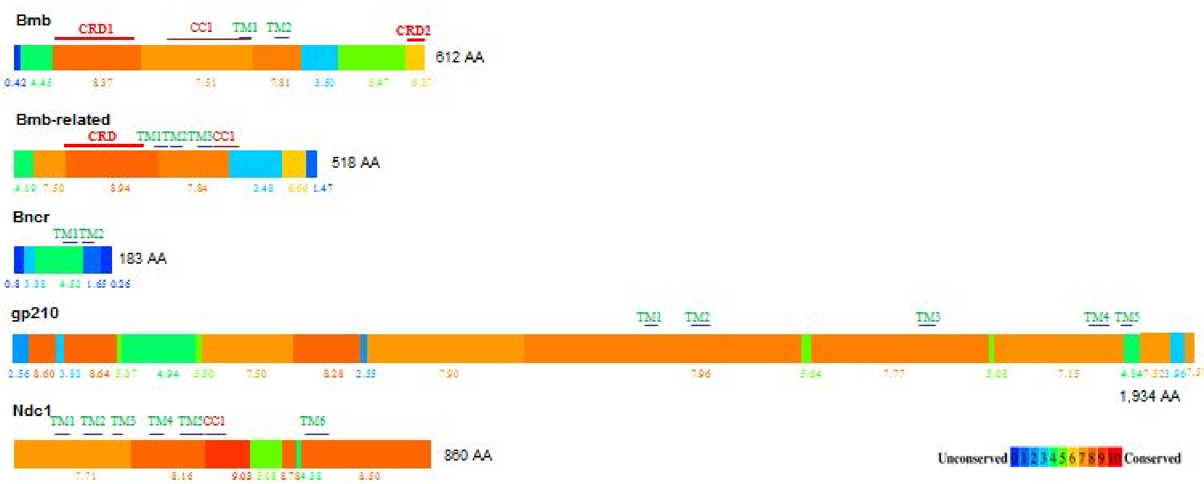
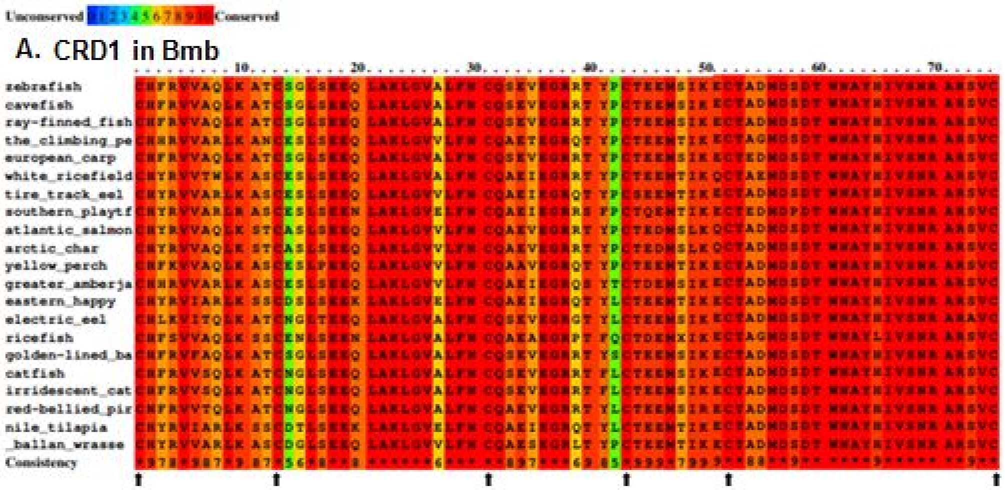


Figure 3: Conserved regions within Bmb, Bmb-related, Bncr, gp210, and Ndc1 proteins. Each protein schematic represents the average conservation of amino acid residues at a given position. For each protein, the top 20 fish homologs closest to zebrafish were chosen and heat maps were determined using the Praline alignment program (http://www.ibi.vu.nl/programs/pralinewww/). Conservation scores (0-10) are indicated (lower = blue; intermediate = green and higher = red). Transmembrane domain (TM) locations were determined by generating hydropathy plots. Coiled-coil (CC) domain locations were determined using (http://cb.csail.mit.edu/cb/multicoil/cgi-bin/multicoil.cgi). Cysteine-rich domains are indicated in Bmb (CRD1 and CRD2) and Bmb-related (CRD).

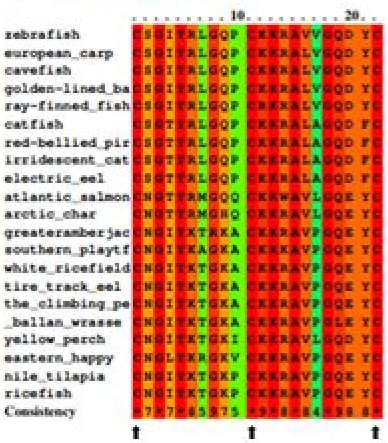
Since the alignments that are run on Praline resulted in long heat maps we condensed the 20 aligned proteins from each category and averaged the sections based on their conservation level. This results in 5 separate schematic composites that show the highly conserved regions as well as the regions that are un-conserved. Each protein differs from the next. (fig.3). The schematics show the regions of their domain locations (tm, cc, and crd). On the number scale located on the lower right corner of the figure, the highest conservation level areas are rated as being a 10 and the un-conserved areas are a 0. Intermediate levels of conservation are located in the green zone. The first two schematics show the similarity of both Bmb and Bmb related. Although they have regions that are un-conserved, the areas that overcome the schematic are on the higher scale of conservation. These two schematics have domains in both their n-terminus as well as the c-terminus. In the n-terminus of Bmb there is a high level of conservation, specifically in the crd's. In the c-terminus it can be seen that there is indeed a lower level of homology. Which is interesting because in Bncr, the entire protein is un-conserved. Note that based on this these two factors alone, it does not make it clear that Bmb is species-specific since there are regions that are highly conserved. Bmb is seen as having two tm domains while Bmb related is mapped out as having three. Their similarity is uncanny but they still have differences that separate them on an evolutionary level.

The tm domains are also located in different areas in both Bmb and Bmb related. Bncr is mapped out as being on the lower level of conservation. The blue and green are labeled in the schematic. It is important to note the numbers that are labeled underneath each schematic. These numbers represent the scale color that shows the conservation levels in each specific protein. The blue mapped out in Bncr links the level of conservation as proof that this specific protein is in fact species-specific. At the end of each schematic lies a label of how long each specific amino acid (aa) sequence is for the protein. Bmb is 612 aa sequences long, Bmb related is 518, Bncr 183, Gp210 1934, and Ndc1 is 860 aa sequences long. The two control npc's (Gp210 and Ndc1) have their tm domains mapped out as well as their cc domains. These two npc's are very conserved with some slight regions that are un-conserved. The high conservation levels in these two proteins show that they can not be species-specific, like that of Bncr. Also, note the regions that are un-conserved in the c-terminus of both Bmb and Bmb related (refer to fig.3).

A. CRD1 in Bmb



B. CRD2 in Bmb



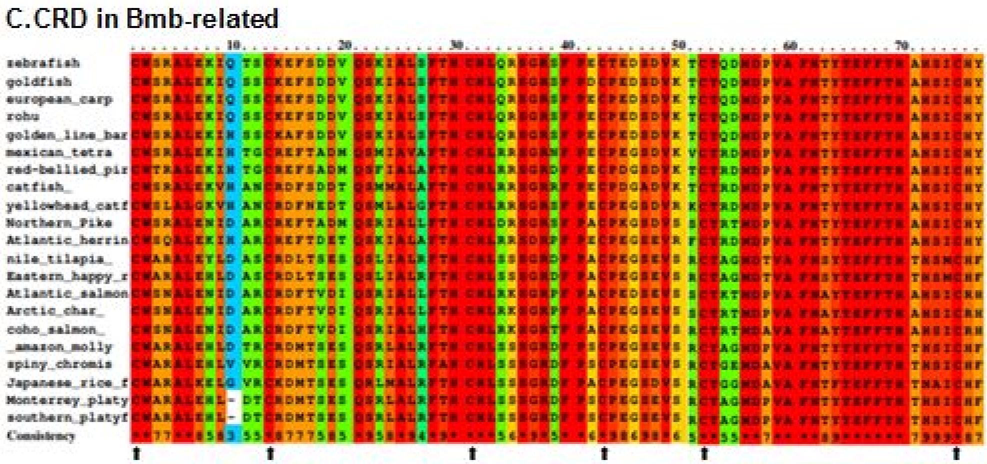


Figure 4. Praline alignments of cysteine-rich domains (CRD) in Bmb and Bmb-related. Bmb N-terminal CRD1 **(A)** and Bmb C-terminal CRD2 **(B)** contain six and three invariant cysteine residues (black arrows), respectively. The CRD in Bmb-related **(C)** also contains six invariant cysteines.

The Bmb and Bmb related conservation figures map out highly conserved regions. They are labeled as being the cysteine-rich domains (CRD), shown in figure 3 and 4 a- c. There are only crd's in Bmb and Bmb related, while the other 3 proteins do not contain these domains. Fig.4 a shows the Bmb homolog created by Praline. There are 6 regions located in the n-terminus that are labeled with black arrows that represent the crd's. The location of these domains are towards the beginning of the protein sequence. In figure 4b, it shows the Bmb homologs c-terminus crd's. There are only 3 crd's located in this region. In figure 4c, this represents the Bmb related protein homolog and this is identical to that of figure 4a which shows the crd's in the same positions as the crd's in the n-terminus of Bmb related. There are no crd's in the c-terminus of Bmb related. This can also be seen in the schematic representation of the homologs (fig. 3a). Although both Bmb and Bmb related are similar, they are not 100% identical to one another, the difference is seen in the scale of conservation (fig4.a, fig4c). The homologs allow us to see the differences in their structure as well as their identical regions in the n-terminus.

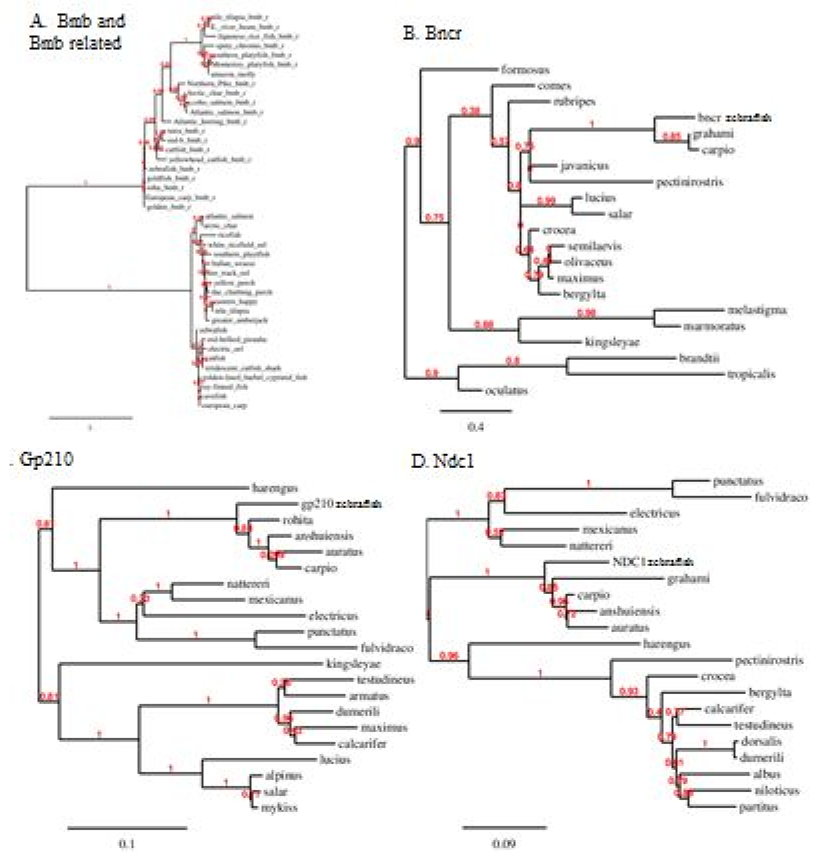


Figure 5. Phylogenetic trees of all 5 protein top 20 species homologs showing linkage through evolution. A. Bmb and bmb related are similar but they break off into two separate clades showing their divergence through evolution. In bmb the Zebrafish is closely linked to the red-bellied piranha, the electric eel, catfish, cyprinid, Ray-finned fish, cavefish and the European carp. With distances ranging from.7 - .9. Bmb related is linked to the goldfish, the Atlantic rohu, Tetra, the European carp and the cyprinid fish. All distances vary from .4-.71. **B**. Bncr is closely related to the cyprinid and the European carp with a distance of .85. **C.** Gp210 is closely related to the Asian rohu with a distance of .83, the cavefish with a distance of 1, the goldfish and the European carp with a distance of .099. **D**. Ndc1 is closely related to the cyprinid fish with a distance of .85, the European carp with a distance of .96, and it is related to both the cavefish and the goldfish with a distance of .72.

In order to show that these fish have significant homology between them we generated phylogenetic trees. Figure 5 a-d shows the lineage of each specific species of fish and how close they are in relation to one another. In fig 5a, there is an obvious link between both the Bmb and Bmb related clade but it diverges and splits into two separate clades. Even though they are similar to one another, they do not intermingle with one another.

They are related to an extent but they have diverged over time as seen in the figure with the lines separating each specific species of fish. The Zebrafish is closely linked to the red-bellied piranha, the electric eel, catfish, cyprind, Ray-finned fish, cavefish and the European carp. The similar fish have distances ranging from.7 - .9. Bmb related is linked to the goldfish, the Atlantic rohu,Tetra, the European carp and the cyprinid fish. All distances vary from .4-.71.Figure 5b shows how Bncr is closely related to the cyprinid and the European carp with a distance of .85. The Gp210 tree (figure 5 c) is closely related to the Asian rohu with a distance of .83, the Mexican cavefish with a distance of 1, the goldfish and the European carp with a distance of .099.and Ndc1 (figure 5 d) is closely related to the cyprinid with a distance of .85, the European carp with a distance of .96, and it is related to both the cavefish and the goldfish with a distance of .72. All of these differences (refer to figure 5) are the size how far the fish are when comparing their lineage to one another.

Throughout evolutionary time fish have diverged from their initial ancestry and many factors can be attributed to this. Phylogenies are useful for organizing knowledge of biological diversity, for structuring classifications, and for providing insight into events that occurred during evolution. Individual populations are bound to be isolated for some period of time. However, on an evolutionary timescale, migration will also occur. This gene flow between populations has the effect of turning a population into a single species lineage. During evolution lineages also tend to split. This occurs when populations or groups of populations become genetically isolated from one another, and if the isolated populations remain separate, they will start evolving differences from one another. This in the end will result in a drift of similar species that act differently because of natural selection (Baum, D., 2008). As a consequence of this genetic isolation, the lineages will evolve separately, becoming more and more different over time. This can be seen in fig 5 a, with Bmb and Bmb related being similar to one another they still break off into clades. If they remain apart for long periods, enough physiological and behavioral differences may evolve to result in reproductive isolation. Keep in mind that the lines of a tree represent evolutionary lineages — and evolutionary lineages do not have any true position or shape. So they will inevitably change over time. The phylogenetic tree represents how species of fish can be similar to one another, but diverge from each other depending on the contributing factor as well as natural selection.

Discussion:

As evolution continues to occur there can still be a level of conservation that never seems to disappear as time passes by. The evolutionary link between the 42 species of fish tells us that there is a specific conservation that all of the fish will not stray away from. This factor is what links the fish as being similar to one another but they are not 100% identical to one another. What this indicates is that Bmb and Bmb related contain highly conserved regions and un-conserved regions. In contrast, Bncr is un-conserved which is why we think that it is species-specific. It is possible that both Bmb and Bmb related have un-conserved regions as seen in the blue sections of the c-terminus (fig3), but in order for us to test this it needs to be done in a lab setting. The homologs seen in fig.4 show proof and representations of conservation. This may be a link that Bmb can potentially have regions that are species-specific, which will function to that of Bncr. The hypothesis of Bncr being un-conserved is supported (fig.3) with the color scale showing its very low conservation levels.

The importance of this study is to view, compare and discover new ways of interpreting the evolutionary significance of vertebrates cell fusion based on their protein sequence. This study focuses on the model organism Danio rerio (zebrafish). In order to do this the bioinformatics approach is best. In an alignment there are certain domains in each protein structure, and these domains will vary from protein to protein. Domains are extremely distinct functional and/or structural units in a protein. These domains are responsible for a particular function or interaction, and this contributes to the overall role of a protein. Three domains (cysteine rich domain, transmembrane domain and coil-coil domain) are mapped in the protein schematic seen in figure 3. Each domain in the Bmb and Bmb related protein analyzed will have some form of similarity and although the regions will never be 100% similar they will have conserved regions and un-conserved regions. Cysteine-rich domains are an important category that involve the high conservation levels in the regions of an aa sequence. They are important to map and point out because they are crucial in the structure of a particular protein. In this case they are extremely important for the structure and function of the Bmb protein. A group of cysteines would map out the highly structured portion of the protein (fig 4a-c). They are also important because they form disulfide bonds; which is a single covalent bond between the sulfur atoms to two amino acids called cysteine. The Bmb homology domain is only found in Bmb and this is a specific domain that is highly conserved making it the extremely important for protein structure. Note that this domain also withstands change which strengthens its structure. The coil-coil domains are important for protein-protein interactions. It may be possible that Bmb can interact with itself, or it may interact with a different protein, such as Bmb related.

This study aims to find the levels of the Bmb protein that may be species specific on the region of the nuclear envelope. We know from prior studies done by Herberg et al which involved using species-specific Bncr to cross-species fertilization between zebrafish and medaka (Japanese rice fish), which are two different species of fish that have been diverged for over 200 million years. This study came to the conclusion that Bncr mediates sperm-egg binding and is needed for sperm entry into the egg (Herberg, S et al., 2018). Further proving that some proteins that are un-conserved may function in being species-specific. In this study Bmb and Bmb related contain both highly conserved regions as well as non-conserved regions. In order to test this the next study would have to involve testing diverged species in order to see if their un-conserved region is still able to fuse when excised and swapped into another fish. If the transformation of that species is able to go through the process of the fusion of the plasma membrane which leads to a one-cell zygote then that particular fish can be linked as being species-specific. This is a forced artificial excising of the fusion protein (Bncr). In regards to Bncr being the gatekeeper, this tactic will change that gatekeeper by forcing it to fuse using another fishes fusion protein. In the end result, this information can be interpreted in a few ways. Bmb may not have a role in species-specific fertilization, but it may have multiple roles of functioning within the protein. In order to test this hypothesis we can test this in the lab using similar tactics that were done for the study of the Bncr protein. Excising Bncr from species A and then placing it into species B, this forced species B to undergo plasma membrane fusion which led to fertilization. Likewise, we would like to do this focusing on the species of zebrafish and the fusing of the Bmb protein being excised. Focusing on the domains seen in fig 3 and 4, we would potentially use two species of fish in the same manner as the Bncr study. We would then excise the un-conserved regions to potentially see if the zebrafish females can cross-breed with a diverged fish such as medaka.

Bibliography:

Abrams, E. W., Zhang, H., Marlow, F. L., Kapp, L., Lu, S., & Mullins, M. C. (2012). Dynamic Assembly of Brambleberry Mediates Nuclear Envelope Fusion during Early Development. Cell, 150(3), 521–532.

Baum, D. (2008) Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups. *Nature Education* 1(1):190

"BLAST: Basic Local Alignment Search Tool." National Center for Biotechnology Information. https://blast.ncbi.nlm.nih.gov/Blast.cgi.

Dun, Matt & Mitchell, Lisa & Aitken, Robert & Nixon, Brett. (2010). Sperm–Zona Pellucida Interaction: Molecular Mechanisms and the Potential for Contraceptive Intervention. Handbook of experimental pharmacology. 198. 139-78. 10.1007/978-3-642-02062-9_9.

European Bioinformatics InstituteProtein Information ResourceSIB Swiss Institute of Bioinformatics. "Nucleoporin NDC1." European Bioinformatics InstituteProtein Information ResourceSIB Swiss Institute of Bioinformatics. May 08, 2019. https://www.uniprot.org/uniprot/Q8VCB1.

Ganguly A, Bukovsky A, Sharma RK, Bansal P, Bhandari B & Gupta SK. (2010). In humans, zona pellucida glycoprotein-1 binds to spermatozoa and induces acrosomal exocytosis. *Hum. Reprod. , 25*, 1643-56.

Herberg, S., Gert, K. R., Schleiffer, A., & Pauli, A. (2018). The Ly6/uPAR protein Bouncer is necessary and sufficient for species-specific fertilization. *Science*, 361(6406), 1029–1033.

Gottschling, Irmia R., (2019), "Understanding Ovulation." *American Pregnancy Association*, doi:10.3897/bdj.4.e7720.figure2f.

"Infertility." Mayo Clinic. March 08, 2018. https://www.mayoclinic.org/diseases-conditions/infertility/symptoms-causes/syc-20354317.

K. Gupta, Satish & Bhandari, Beena & Shrestha, Abhinav & Biswal, Bichitra & Palaniappan, Chetna & Malhotra, Sudha & Gupta, Neha. (2012). Mammalian zona pellucida glycoproteins: Structure and function during fertilization. Cell and tissue research. 349. 665-78.

Herberg, S., Gert, K. R., Schleiffer, A., & Pauli, A. (2018). The Ly6/uPAR protein Bouncer is necessary and sufficient for species-specific fertilization. Science, 361(6406), 1029–1033. doi:10.1126/science.aat7113
Lehman, R., (2018), Matchmaking molecule for egg and sperm, *Science*, Vol 361, Issue 6406, pp. 974-975

Monroy, Alberto. (November 22, 2018). "Fertilization." *Reproduction*, doi:10.3897/bdj.4.e7720.figure2f.

Mold, D. E., Dinitz, A. E., & Sambandan, D. R. (2009). *Regulation of Zebrafish Zona Pellucida Gene Activity in Developing Oocytes 1. Biology of Reproduction, 81(1), 101–110.*doi:10.1095/biolreprod.108.071720

"Ovulation - Understanding Ovulation Cycles." American Pregnancy Association. May 14, 2019. https://americanpregnancy.org/getting-pregnant/understanding-ovulation/

Simossis, V. A., & Heringa, J. (2005). *PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. Nucleic Acids Research, 33(Web Server), W289–W294.*

Schoft, Vera K., Ariane J. Beauvais, Carmen Lang, Andreas Gajewski, Kristina Prüfert, Christoph Winkler, Marie-Andrée Akimenko, Micheline Paulin Levasseur, and Georg

Krohne. (March 4, 2003). "Research Article." *The Lamina-associated Polypeptide 2 (LAP2) Isoforms β, γ and ω of Zebrafish: Developmental Expression and Behavior during the Cell Cycle* 10, no. 7: 2505-517.

"What Are Some Possible Causes of Female Infertility?" Eunice Kennedy Shriver National Institute of Child Health and Human Development. January 31, 2017. https://www.nichd.nih.gov/health/topics/infertility/conditioninfo/causes/causes-female.

Xu, S, Powers, A. M. (2009), "Nuclear Pore Proteins", *Elsevier*, vol 20, issue 5, p.620-630