HHS Public Access

Author manuscript

Immunogenetics. Author manuscript; available in PMC 2024 December 03.

Published in final edited form as:

Immunogenetics. 2015 October ; 67(10): 605–623. doi:10.1007/s00251-015-0870-1.

Molecular Evolution and *in vitro* Characterization of *Botryllus histocompatibility factor*

Daryl A. Taketa,
Department of Molecular, Cellular and Developmental Biology, University of California – Santa Barbara, Santa Barbara, CA 93106, USA

Marie L. Nydam,
Division of Science and Mathematics, Centre College, Danville, KY 40422, USA

Adam D. Langenbacher,
Department of Molecular, Cellular and Developmental Biology, University of California – Santa Barbara, Santa Barbara, CA 93106, USA

Delany Rodriguez,
Department of Molecular, Cellular and Developmental Biology, University of California – Santa Barbara, Santa Barbara, CA 93106, USA

Erin Sanders,
Department of Molecular, Cellular and Developmental Biology, University of California – Santa, Barbara Santa Barbara, CA 93106, USA

Anthony W. De Tomaso
Department of Molecular, Cellular and Developmental Biology, University of California – Santa, Barbara Santa Barbara, CA 93106, USA

Abstract

Botryllus schlosseri is a colonial ascidian with a natural ability to anastomose with another colony to form a vascular and hematopoietic chimera. In order to fuse, two individuals must share at least one allele at the highly polymorphic *fuhc* locus. Otherwise, a blood-based inflammatory response will occur resulting in a melanin scar at the sites of interaction. The single-locus genetic control of allorecognition makes *Botryllus schlosseri* an attractive model to study the underlying molecular mechanisms. Over the past decade, several candidate genes involved in allorecognition have been identified, but how they ultimately contribute to allorecognition outcome remains poorly understood. Here we report our initial molecular characterization of a recently identified candidate allodeterminant called *Botryllus histocompatibility factor (bhf)*. *bhf*, both on a DNA and protein level, is the least polymorphic protein in the *fuhc* locus studied so far and, unlike other known allorecognition determinants, does not appear to be under any form of balancing or directional selection. Additionally, we identified a second isoform through mRNA-Seq and an EST

Corresponding Author: A. W. De Tomaso, anthony.detomaso@lifesci.ucsb.edu, Tel: +1 805 893 7276.
Present address: E. Sanders, Department of Developmental Biology, Stanford University, Stanford, CA 94505

Conflict of Interest
The authors declare that they have no conflict of interest.

assembly library which is missing exon 3, resulting in a C-terminally truncated form. We report via whole-mount FISH that a subset of cells co-express *bhf* and *cfuhc*^{sec}. Finally, we observed BHF's localization in HEK293T at the cytoplasmic side of the plasma membrane in addition to the nucleus via a nuclear localization signal (NLS). Given the localization data thus far, we hypothesize that BHF may function as a scaffolding protein in a complex with other *Botryllus* proteins, rather than functioning as an allorecognition determinant.

Keywords

Botryllus schlosseri ; Allorecognition; *Botryllus* Histocompatibility Factor (BHF); Cellular localization; Selection

Introduction

Botryllus schlosseri is a colonial marine invertebrate that belongs to the phylum Chordata, subphylum Tunicata. Tunicates have been identified as the closest living relatives to the vertebrates (Delsuc et al. 2006, 2008), making *B. schlosseri* an interesting model to study fundamental biological processes like allorecognition: the ability for an organism to distinguish self from non-self. Allorecognition has been well-studied in the context of the major histocompatibility complex (MHC) based immunity in vertebrates. However, the evolutionary origins of the MHC and the mechanisms that regulate recognition, education, and tolerance remain a mystery. Burnet (1971) postulated that invertebrate metazoans hold the answer to these questions as allorecognition responses have been observed across the tree of life (reviewed in Rosengarten and Nicotra 2011).

B. schlosseri displays a natural transplantation ability that can allow an individual to anastomose its extra-corporeal vasculature system with that of another individual (fusion), creating a vascular and hematopoietic chimera. This allorecognition system is genetically controlled by a single, highly polymorphic *fusion/histocompatibility* (*fuhc*) locus (Scofield et al. 1982; Weissman et al. 1990). Fusion occurs when individuals share one or both *fuhc* alleles. When no alleles are shared, the two individuals will undergo a blood-based inflammatory response that results in a melanin scar at the sites of interaction (rejection). This parallels the mechanism of allorecognition in Natural Killer cells' "missing self recognition" where the target cell is destroyed if no self-markers are present (reviewed in Long et al. 2013). Initially, Scofield *et al.* (1982) proposed that the *fuhc* locus could be the evolutionary link to the MHC. Later studies (De Tomaso et al. 2005; Nydam et al. 2013; Voskoboynik et al. 2013B) showed that *B. schlosseri* does not possess an ancestral form of the MHC based on sequence homology. Additionally, jawless-vertebrates utilize a different system (variable lymphocyte receptors – VLRs) that is independent of the MHC-based immunity based on sequence homology, but acts similarly to T or B cell receptors (reviewed in Boehm et al. 2012). One hypothesis is that the cell-surface components of allorecognition evolved independently, but the mechanisms underlying recognition (e.g., education and tolerance) are conserved (reviewed in De Tomaso 2014). If this is the case, the allodeterminant in *B. schlosseri* should have similar genetic properties to the MHC, such

as the degree of polymorphisms maintained by balancing or directional selection within a population.

Within the *fuhc* locus, several candidate genes are involved in the allorecognition reaction (De Tomaso et al. 2005; Nyholm et al. 2006; McKitrick et al. 2011; Voskoboynik et al. 2013B; Nydam et al. 2013; reviewed in Taketa and De Tomaso 2015). Amongst these genes, *cfuhc*^{sec} (secreted) and *cfuhc*^{tm} (transmembrane) are the most polymorphic genes identified so far (Nydam et al. 2012; Voskoboynik et al. 2013B) making one or both of these genes a likely candidate for a self-recognition molecule. However, there has been controversy identifying the true allodeterminant in *B. schlosseri* (De Tomaso et al. 2005; Rinkevich et al. 2012; Voskoboynik et al. 2013B; Litman and Dishaw 2013; Nydam et al. 2013; Taketa and De Tomaso 2015) primarily due to the lack of sufficient functional evidence. A recent candidate in the *fuhc* locus, called the *Botryllus histocompatibility factor* (*bhf*), was identified ~62 kb upstream of the *cfuhc*^{sec} gene (Voskoboynik et al. 2013B). In this study, the authors reported *bhf* to be a more consistent predictor of fusibility outcome than either *candidate fusion/histocompatibility* secreted (*cfuhc*^{sec}) or the transmembrane form (*cfuhc*^{tm}). *bhf* encodes a 252 amino acid (aa) protein that lacks predicted domains, but is highly charged and partially disordered. Through morpholino-mediated knockdown of *bhf*, a fusion or rejection was prevented between two colonies resulting in a “no reaction” phenotype implicating a role in allorecognition (Voskoboynik et al. 2013B). Collectively, this has led to the hypothesis that *bhf* is *the* allodeterminant, but the mechanism remains elusive considering it lacks any predicted or conserved features in signaling or cell surface presentation that would explain the knockdown phenotypes (reviewed in Taketa and De Tomaso, 2015).

In this study, we report that *bhf* has the lowest nucleotide and amino acid diversity of all the candidate allorecognition proteins in the *fuhc* locus and shows no evidence of balancing or directional selection (in contrast to *cfuhc*^{tm}, *cfuhc*^{sec}, *fester* and *hsp40-1*). We also identified a truncated BHF isoform that lacks a putative nuclear localization signal (NLS) at its carboxyl terminal end. We re-evaluated *bhf* expression localization and discovered a subset of *bhf* (*isoform 1* and *2*) positive cells were also expressing *cfuhc*^{sec}. Finally, utilizing a mammalian tissue culture expression system, we have found that BHF is plasma membrane associated at the cytoplasmic side and has the potential to localize to the nucleus.

Materials and Methods

Polymorphism Analysis

RNA isolation, cDNA synthesis, bhf amplification and sequencing—Maternal colonies were collected from seven locations around Santa Barbara harbor in April-September 2011 and were allowed to release larvae onto microscope slides. Metamorphosed larvae (oozoids), all from different mothers, were transferred in groups of five to new microscope slides. These slides were taken to the Santa Barbara harbor one week after the oozoids were transferred, and were submerged 0.5 meters below the surface of the water. After 1–2 months, colonies were transported back to the laboratory. A single system from each colony was frozen in liquid nitrogen and immediately stored at −80°C. No colonies used in this study were observed to have fused with other colonies on the same microscope

slide. However, they likely fused with larvae present in the seawater, or with larvae released from colonies on the same or nearby microscope slides.

Sixteen systems were selected, representing all seven Santa Barbara harbor locations, for *bhf* sequencing. Total RNA was extracted from frozen tissue using the NucleoSpin® RNA Kit (Macherey-Nagel, 740955). This RNA was used to synthesize single-stranded cDNA using SuperScript III reverse transcriptase (Life Technologies, 18084-044) and an oligo (dT) primer. PCR amplification was performed with Phusion Polymerase (Thermo Fisher Scientific, F-530L). The forward primer and reverse primer used are designed against the untranslated region (UTR) of *bhf* (Table 1). Cycling conditions were 98°C for 1 minute, 35× (98°C for 30 seconds, 58°C for 30 seconds, 72°C for 30 seconds), 72°C for 5 minutes.

PCR products were cloned using the pGEM®-T-Easy kit (Promega, A1360) and eight clones were sequenced. Colony PCR products were incubated with 0.25μl each of Exonuclease I and Shrimp Antarctic Phosphatase at 37°C for 30 min, followed by 90°C for 10 min prior to sequencing. Purified PCR products were sequenced with a Big Dye Terminator Cycle sequencing kit and a 96 capillary 3730×l DNA Analyzer (Applied Biosystems) at the UC Berkeley Sequencing Facility. All sequences have been submitted to GenBank (Accession Numbers KR911864-KR911892). Sequences were edited, trimmed and aligned with SeqMan Pro (DNASTAR, Inc., Madison, WI).

Amino acid diversity across BHF—The program DIVAA (Rodi et al. 2004) was employed to calculate the amino acid variation across BHF for the 29 alleles recovered from 16 individuals. A diversity value of 1 for a particular position means that any amino acid is as likely as any other to be present at that position. A diversity value of 0.05 is consistent with 100% conservation at that site (1/20 possible amino acids). Amino acid variation was visualized using R 3.1.1 (R Core Team 2014).

Tests of selection: ω—The ω (dn/ds) value and associated 95% highest posterior density (HPD) region across BHF was estimated using the program omegaMap0.5 (Wilson and McVean 2006). omegaMap0.5 runs were carried out using the resources of the Computational Biology Service Unit at Cornell University. 250,000 iterations were chosen for each run, with thinning set to 1,000. An improper inverse distribution for μ (rate of synonymous transversion), and κ (transition-transversion ratio) and an inverse distribution for ω (selection parameter) and ρ (recombination rate) were used. Initial parameter values for μ and κ were 0.1, and 3.0, respectively. Ω and ρ priors were set between 0.01 and 100. An independent model was used for ω, so that ω values were allowed to vary across sites. The number of iterations discarded as burn-in varied across runs, but was determined by plotting the traces of μ and κ; iterations affected by the starting value of the parameter were discarded. Two independent runs were conducted. These two runs were combined, after it was determined that the mean and 95% HPD regions for each parameter in the two runs matched closely. The posterior probability of selection per codon was also determined. Mann-Whitney U tests in R 3.1.1 (R Core Team 2014) were performed to determine if any of the three exons had higher selection statistic values than the rest of the transcript.

Tests of selection: Tajima 's D and Fu and Li's D* and F*—The summary statistics θ, π, number of haplotypes and haplotype diversity were calculated in DnaSP 5.10.1 (Librado and Rozas 2009). The Tajima's D value (Tajima 1989) and Fu and Li's D* and F* values (Fu and Li 1993) based on total number of mutations were calculated in DnaSP 5.10.1 (Librado and Rozas 2009). Statistical significance of D, D* and F* were determined using 10,000 coalescent simulations in DnaSP 5.10.1 (Librado and Rozas 2009). Coalescent simulations were based on θ. An estimate of per gene recombination was made in DnaSP 5.10.01 (Librado and Rozas 2009) and was then imported into the simulations.

Protein expression in mammalian cells

Animals, cDNA synthesis, and cloning—*Botryllus schlosseri* colonies were collected from the harbor in Santa Barbara, CA with their progenies spawned and cultured under laboratory conditions as previously described (Braden et al. 2014). Lab-reared colonies were collected, homogenized to a frozen powder with a mortar and pestle on dry ice with liquid nitrogen, and then stored at −80°C.

Total RNA was extracted from frozen colonies with NucleoSpin® RNA (Macherey-Nagel, 740955). Samples were concentrated with NucleoSpin® RNA Clean-up XS (Macherey-Nagel, 740903) as needed. For downstream cloning into mammalian expression vector, mRNA was purified from 4 μg of total RNA with a magnetic mRNA isolation kit (New England Biolabs, S1550S). Reverse transcription was conducted with M-MLV (New England Biolabs, M0253S) using random hexamer (Life Technologies, 48190-011) according to manufacturer's instruction.

cDNA generated from mRNA was used to amplify and isolate *bhf*. Full length *bhf* for isoform 1 or 2 were amplified using custom primers (Integrated DNA Technologies) designed against the untranslated regions (UTR) flanking the start/stop codons (Table 1) with Advantage® 2 Polymerase Mix (Clontech, 639201) following the recommended manufacturer's protocol. Amplicons were ligated into pGEM®-T-Easy and sequenced through either Genscript or the UC Berkeley Sequencing Facility. Mammalian expression constructs for full length and truncated forms of *bhf* (isoform 1) were generated with custom primers that included restriction sites for pmCherry-N1 (Clontech, 632523) (NheI & XhoI) or pmCherry-C1 (Clontech, 632524) (XhoI & BamHI). Mammalian expression plasmids for isoform 2 were also cloned using the same method: pmCherry-N1 (NheI & XhoI) and pmCherry-C1 (XhoI & BamHI). BHF (isoform 1) with a FLAG tag was generated through custom primers that attached the FLAG tag on the 3' end and was cloned into pCMV-Tag3B (Agilent, 211173-51) through a NotI (5') and a XhoI (3') restriction site (Table 1). BHF (isoform 1) cysteine mutant was generated through Integrated DNA Technologies gBlocks® Gene Fragments synthesis with Cys20, Cys33, and Cys40 mutated into serines with a XhoI (5') and a BamHI (3') restriction site added and cloned into pmCherry-C1.

Tissue Culture—Human embryonic kidney (HEK) 293T cells were cultured in DMEM high glucose with L-glutamine and sodium pyruvate (GE Healthcare Hyclone, SH30243FS) supplemented with 10% fetal bovine serum (Atlanta Biologicals, F-0500-A), 100 units/ml penicillin and 100 μg/ml streptomycin (GE Healthcare, SV30010) at 37°C in 5% CO_2.

Cells were seeded on nitric acid washed glass coverslips coated with poly-L-lysine (Electron Microscopy Sciences, 19321-B) and collagen type I (Corning, 354236) in a 6-well plate at a confluency of 20 – 30% and were allowed to attach for 24 hours. If applicable, cells were pretreated with 100 μM of 2-bromopalmitate (Sigma-Aldrich, 21604) in DMSO for 30 minutes to inhibit any post-translation palmitoylation before transfection. Cells were transfected with the plasmids using Lipofectamine 2000 (Life Technologies, 11668-027) as per manufacturer's instructions. After 24 to 48 hours post transfection, cells were treated with 100 μg/ml cycloheximide (Sigma-Aldrich, C-0934) for at least 2 hours when indicated. Transient expressing cells were fixed in 4% formaldehyde (Thermo Fisher Scientific, BP531-500) in 1× PBS (pH 7.4) at room temperature for 20 minutes or fixed with 4% formaldehyde methanol free (Electron Microscope Sciences, 15710) in 1× PBS on ice for 20 minutes for cell surface staining. Cells were either mounted in VectaShield with DAPI (Vector Laboratories, H-1200) or immunostained (described below).

Immunostaining—Fixed cells were permeabilized with 0.25% Triton X-100 (if applicable) for 15 minutes followed by blocking in 2% (w/v) BSA in 1×PBS for 30 minutes at room temperature. Primary antibodies were incubated overnight at 4°C in 1×PBS/1% (w/v) BSA as follows: rat α-mCherry (Life Technologies, M11217) was diluted 1:1000; rabbit α-GPR50 (Cell Signaling Technology, 14032) was diluted 1:400; mouse α-Lamin A/C (Cell Signaling Technology, 4777) was diluted 1:200; and mouse α-Nucleolin (Life Technologies, 39-6400) was diluted 1:100. Cells were then incubated with their respective secondary antibody: Alexa Fluor® 488 goat α-rat (Life Technologies, A-11006), goat α-rabbit (Life Technologies, A-11008), or goat α-mouse (Life Technologies, A-21121) diluted 1:1000 in 1×PBS/1% BSA at room temperature for 1 hour. After the last wash step in 1×PBS, cells were mounted in VectaShield with DAPI and sealed with nail polish.

Protease K Protection Assay, Subcellular Fractionation, and Western Blot—Transient expressing cells in a 24 well plate were washed with cold 1× PBS on ice twice. Whole cells were treated in the presence or absence of 10 μg/ml of Protease K (Roche, 03115836001) in the presence or absence of 1% NP-40 in 1× PBS for 30 minutes on ice. Protease K digestion was inhibited by the equal volume addition of 2× Laemmli buffer supplemented with 2 mM PMSF. Cell lysates were transferred to microfuge tubes then briefly sonicated. Samples were heated at 70°C for 10 minutes then centrifuged at 17,000 × g for 10 minutes. Equal volumes of each sample were loaded on a 12% SDS-PAGE gel then transferred to a nitrocellulose membrane.

Subcellular fractions of transient expression mCherry or mCherry:BHF cells were collected through a subcellular fractionation kit (Life Technologies, 78840) following the manufacturer's instructions. Equal volumes from each fraction was loaded on a 12% SDS-PAGE gel then transferred to a nitrocellulose membrane.

Membranes were blocked with 5% (w/v) non-fat dried milk in PBST (0.1% Tween-20) for 1 hour then incubated with rat α-mCherry, rabbit α-EGFR (Cell Signaling Technology, 4267), or rabbit α-HSP90 (Cell Signaling Technology, 4874) antibody diluted 1:1000 in blocking solution overnight at 4°C. Washed membranes were incubated with Dylight 680 conjugated goat α-rat antibody (Life Technologies, SA5-10022) or Dylight 800 conjugated

goat α-rabbit (Life Technologies, SA5-35571) diluted 1:5000 in 1× PBST and imaged on a Li-Cor Odyssey system.

mRNA-Seq

Using mRNA-Seq data (GEO accession# GSE62112) from Rodriguez *et al.* (2014), reads were mapped to the draft genome (Voskoboynik et al. 2013A) merged with genomic sequencing from De Tomaso and Weissman (2003B). We employed the workflow for Cufflinks (version 2.1.1) as described in Trapnell *et al.* (2010) using a custom annotation file for the *fuhc* locus. Mapped reads were visualized using IGV Version 2.3.34 (Robinson et al. 2011; Thorvaldsdóttir et al. 2012) against the genomic sequence. Additionally, the mRNA-Seq data was also mapped to a public *Botryllus* EST assembly (http://octopus.obs-vlfr.fr/public/botryllus/blast_botryllus.php, database: "Bot_asmb assembly 04.05.2011, A. Gracey") and followed the workflow for DESeq (version 1.10.1) analysis (Anders and Huber 2010) with the same computer setup as previously described in Rodriguez *et al.* (2014). mRNA-Seq samples from 2 fertile and 2 infertile animals were pooled together for each blastogenic stage.

Whole mount fluorescent *in situ* hybridization

Whole mount fluorescent *in situ* hybridization (FISH) was conducted as previously described (Langenbacher et al. 2015). Briefly, young adult colonies were anesthetized with 0.02% (w/v) Tricaine (TCI, T0941) in seawater for 10 minutes then fixed in 4% formaldehyde in 0.5 M NaCl, 0.1 MOPS pH 7.5 at room temperature with agitation for 3 hours. Samples were dehydrated with methanol and bleached in 6% H_2O_2 in methanol under direct light. Once bleached, samples were stepwise rehydrated in PBST then permeabilized with 10 μg/ml protease K (Roche, 03115879001) for 30 minutes at room temperature. Samples were then fixed again in 4% formaldehyde in PBST for 20 minutes at room temperature, washed with PBST, then incubated in hybridization buffer (65% formamide, 5× SSC, 1× Denhardt's solution, 0.1% Tween-20, 5 mg/ml torula yeast, 50 μg/ml heparin) for 4–6 hours at 58–65°C. Afterwards, samples were incubated digoxigenin (DIG)-labeled riboprobes diluted in hybridization buffer overnight at 58–65°C. Probes were diluted to a final concentration ranging between 400 – 1200 pg/μl. Unbound probe was washed away, and samples were then incubated in blocking buffer (PBST, 5% (v/v) heat-inactivated horse serum, 2 mg/ml (w/v) bovine serum albumin) for 4 hours at room temperature. Incubation with HRP-conjugated anti-digoxigenin antibody (Roche, 11207733910) diluted 1:1000 in blocking buffer was performed overnight at 4 °C. Samples were washed with PBST + 2 mg/ml (w/v) bovine serum albumin, then PBS, and digoxigenin-labeled probes were detected by a fluorophore deposition using the TSA Plus System (Perkin Elmer, NEL753001KT). For double FISH, a dinitrophenol (DNP)-labeled probe was co-incubated with the DIG-labeled probe. After the HRP-conjugated anti-DIG antibody was inactivated, DNP-labeled probe was detected with a HRP-conjugated anti-DNP antibody (Perkin Elmer, FP1129) at a 1:200 dilution followed by a second fluorphore deposition using the TSA Plus System. Samples were counterstained with DAPI then flat-mounted with Vectashield (Vector Labs, H-1000).

Imaging

All fluorescent images were acquired on an Olympus FLV1000S Spectral Laser Scanning Confocal microscope either under a 40× or 60× oil immersion objective, optical slices were taken at 1 μm steps for tissue culture cells and 1.5 μm steps for FISH. Cell sizes from FISH images were measured using FIJI software (Schindelin et al. 2012).

Results

BHF does not have genetic characteristics of a typical allodeterminant

The defining characteristic of allodeterminants is a high level of polymorphisms maintained under a directional or balancing selection (Grosberg 1988). We investigated if *bhf* shares this common characteristic through several statistical analyses based on the population within the Santa Barbara harbor: DIVAA, ω, Tajima's D, and Fu and Li's D* and F*.

DIVAA reveals low amino acid diversity across BHF—The amino acid diversity across BHF for each amino acid position is presented in Fig. 1a. Of the 252 amino acid sites, 237 amino acids (94%) are entirely conserved (Fig. 1b); there is no amino acid variation across 29 sequences from 16 individuals at those sites. The average amino acid diversity across the entire protein is 0.051, and 0.071 for the 6% of sites that are not conserved. For comparison, we randomly picked 29 cFuhc^{sec} sequences from our database of Santa Barbara harbor sequences. We ran a DIVAA analysis on this dataset, and found that 394 amino acid sites (83%) are entirely conserved. The average amino acid diversity across the entire protein is 0.055, and 0.079 for the 17% of sites that are not conserved.

bhf is unlikely under a balancing or directional selection via ω test—In *bhf*, there are five codons that showed >95% posterior probability of balancing or directional selection (ω < 1): Codons 39, 59, 102, 106, 194. Codons 39, 59, 102 and 106 are in exon 1, and Codon 194 is in exon 2. 2% of *bhf's* codons showed evidence of balancing or directional selection in the Santa Barbara population, which is a substantially lower percentage than any of the other genes in the *fuhc* locus. For the Santa Barbara population only, 5.6% of *cfuhc^{sec}* codons show evidence of balancing or directional selection, 4.1% of *cfuhc^{tm}* codons, and 4.9% of *hsp40-l* codons. In contrast to other genes in the *fuhc* locus, *bhf* does not appear to be experiencing balancing or directional selection. Although there are no predicted domains in the BHF protein (Voskoboynik et al. 2013B), we can test whether exon 1, 2 or 3 has significantly higher ω values than the rest of the *bhf* transcript. Neither exon 1, 2 or 3 considered alone has significantly higher ω values than the other two exons considered together. Exon 1 vs. Exon 2/3: W value = 4790, p value = 0.91, Exon 2 vs. Exon1/3: W = 7754, p value = 0.28, Exon 3 vs. Exon 1/2: W value = 4453, p value = 0.24.

Tajima's D and Fu and Li's D* and F* test of selection recapitulates the ω test—Table 2 shows the number of sequences, number of haplotypes, haplotype diversity, π and θ for the Santa Barbara population for *bhf, cfuhc, ^{sec}, cfuhc^{tm}* and *hsp40-l*. Although *bhf* has approximately double the number of sequences, *bhf* has the lowest haplotype diversity, π, and θ_w values of all four genes.

Results from Tajima's D and Fu and Li's D* and F* tests are shown in Table 3. D, D* and F* values for *bhf* are all statistically equivalent to zero, when all sites are considered together, when nonsynonymous sites are considered alone, and when synonymous sites are considered alone. These results provide no evidence for selection acting on *bhf*. In comparison, *cfuhc*^{sec} shows evidence of selection in 2/6 populations when all sites are considered, 2/6 populations when nonsynonymous sites are considered, and 3/6 populations when synonymous sites are considered (Nydam et al. 2012). *cfuhc*^{tm} shows evidence of selection in 3/6 populations when nonsynonymous sites are considered (Nydam et al. 2012).

The bhf gene produces two transcripts

Despite the low levels of polymorphisms and unlikely probability of selection, we next investigated the expression levels of *bhf* using our previously published mRNA-Seq database. We used the mRNA-Seq reads from Rodriguez *et al.* (2014) and mapped them to the genomic sequence of *bhf*. Additionally, BLASTN (version 2.2.29+; Zhang et al. 2000) hits from a public *Botryllus* EST assembly (http://octopus.obs-vlfr.fr/public/botryllus/blast_botryllus.php, database: “Bot_asmb assembly 04.05.2011, A. Gracey”) were aligned as well (Fig. 2a). From this composite, there were a noticeable number of aligned reads downstream of the annotated *bhf* exon 2 where Cufflinks (Trapnell et al. 2010) identified a new isoform. In agreement with the Cufflink results, there were BLASTN hits from the EST assembly that aligned to this extend region in addition to the earlier portion of *bhf*. This new isoform contained an extension of exon 2 in place of exon 3, and resulted in a truncated form of the original reported sequence in Voskoboynik *et al.* (2013B) where only 219 out of 252 amino acids (aa) were present, of which only 6 aa were unique to isoform 2. We amplified this form from mRNA of whole colonies and confirmed the presence of this new isoform via reverse transcription polymerase chain reaction (RT-PCR) (Fig. 2b) and sequencing of the amplicon (GenBank KR911893). This sequence matched to the coding region of contig “Bot_rep_c46042” from the EST library assembly.

We revisited a DESeq analysis from Rodriguez *et al.* (2014) to assess expression levels and graphed the normalized counts that DESeq calculated based on 4 genotypes (2 fertile and 2 infertile) per blastogenic stage for *bhf* isoform 1 and 2 and *cfuhc*^{sec} (Fig. 2c–e). We also conducted a DESeq analysis across the blastogenic stages (Fig. 2c–e) and between ampullae tissue (n=15) and whole colony (n=28) (Fig. 2f) with *cfuhc*^{sec} as a known ampullae tissue enriched gene (Nydam et al. 2013). For both *bhf* isoform 1 (Fig. 2c) and isoform 2 (Fig. 2d), stage D tended to have a higher expression but was not statistically significant compared to any other stage (padj > 0.05; FDR = 0.1) while *cfuhc*^{sec} (Fig. 2e) remained statistically constant. Additionally, when a DESeq was performed between ampullae tissue and whole colony (Fig. 2f), there was no statistically difference for *bhf* isoform 1 (padj = 0.611) meanwhile *bhf* isoform 2 was statistically different (padj = 0.032) and *cfuhc*^{sec} as well (padj = 4.05×10^{-8}) indicating an enrichment of these transcript in the ampullae tissue.

bhf expression patterns differ between juvenile and adult colonies

We reinvestigated the expression pattern of *bhf* in whole-mount samples utilizing our fluorescent *in situ* hybridization (FISH) protocol as previously described (Langenbacher et al. 2015; Fig. 3a–g’’’). Expression of *bhf* (isoform 1) was limited to a subpopulation of cells

within the ampullae (Fig. 3a), resorbing zooid (stage D; Fig. 3b), and secondary bud (Fig. 3c) of adult colonies (>3 months). We also used a smaller probe corresponding to a 102 aa C-terminal deletion mutant (*bhf CΔ102*) that can hybridize with both isoforms. We observed a similar expression pattern in adults' ampullae (Fig. 3d) as we did with the full length *bhf* probe. In these samples, we did not observe any significant signal localized to the epithelium and blood vessels, in contrast to previous results (Voskoboynik et al. 2013B). In that work, only juvenile (~1 month) colonies were used. We thus performed a double *in situ* utilizing the truncated *bhf CΔ102* and *cfuhc*^{sec} probes on both adults and juveniles (<3 months). We observed expression of *bhf* (isoform 1 or 2) along the epithelium of the ampullae (Fig. 3e") and regions of the blood vessels (Fig. 3f') in juveniles. The adult colonies displayed similar expression patterns as the single FISH (data not shown). Additionally, we observed a subset of blood cells co-expressing *bhf* (isoform 1 or 2) and *cfuhc*^{sec} (white arrows; Fig. 3e–g''') within the ampullae (Fig. 4e–e'''), blood vessels (Fig. 3f–f'''), and near the endostyle (Fig. 3g–g'''). We measured the cell size of the expressing blood cells based on the feret diameter in Fiji analysis software (Schindelin et al. 2012) but no significant differences were observed: *bhf* = 10.32 ± 1.89 μm (mean ferret diameter ± standard deviation; n = 408); *cfuhc*^{sec} = 8.94 ± 1.37 μm (n = 89); *bhf & cfuhc*^{sec} = 10.33 ± 1.87 μm (n = 22). Based on size, these cells were within the ranged of a few blood cells types previously reported (Schlumpberger et al. 1984; Ballarin and Cima 2005): macrophage-like cells, morula, and pigment cells. FISH markers to further identify the specific lineage of these cells are still being developed. Generally, *bhf* expressing cells in juvenile colonies appeared to be in clusters of blood cells compared to sparse *cfuhc*^{sec} expressing cells, but this could be attributed to the observed samples size of the *cfuhc*^{sec} expressing population. For example, in the adult colonies (Fig. 3a–d) *bhf* expressing cells were not observed in similar clusters.

BHF expression in HEK293T is localized to the plasma membrane

There are no antibodies directed against BHF and we do not have the ability to express tagged proteins *in vivo*. Therefore, as a first step to determine subcellular localization of BHF, we expressed a tagged version in HEK293T cells. We have previously shown that *Botryllus* proteins fused to a fluorescent protein (e.g., GFP) expressed in mammalian cells were targeted to expected locations based on predicted domains (e.g., transmembrane domain and signal peptides) (Nyholm et al. 2006; McKitrick et al. 2011; Nydam et al. 2013; unpublished data): Fester, Uncle Fester and cFuhc^{tm} localized to the plasma membrane, while cFuhc^{sec} was in the secretory pathway and secreted into the media. However, it should be noted that in this mammalian overexpression system the subcellular localization may not completely reflect how the protein-of-interests behave in *B. schlosseri*. Nonetheless, overexpression of tagged proteins in mammalian culture systems has been successfully used to characterize proteins-of-interest from other marine organisms lacking *in vivo* expression techniques (Kawai et al. 2003; Kondoh et al. 2003; Sasaki et al. 2009). When we expressed an N-terminally tagged BHF, we found that despite lacking any predicted domains or transmembrane regions, it was localized to the plasma membrane regardless of isoform (Fig. 4a–b) in cycloheximide treated cells. Given that isoform 2 is a truncated version of isoform 1 (discussed above), the mechanism of localization is most likely identical. This implies three possible modes of localization for BHF. First, BHF may be a peripheral membrane protein via an electrostatic interaction (considering how many charged residues were present

throughout the protein) or a post translational modification (e.g., lipidation) to anchor it to the membrane. Second, BHF may be a transmembrane associated protein through a potential protein-protein interaction. Third, BHF may associate with the plasma membrane through its lysine rich region (Lys49 to Lys97) in combination with key hydrophobic amino acids as described in Heo *et al.* (2006). We confirmed that the localization is at the plasma membrane by immunostaining BHF expressing cells with a rabbit α-GPR-50 antibody (Fig. 4c–d); where compared to the mCherry expressed control cells (Fig. 4c), BHF expressed cells showed colocalization (Fig. 4d). We further demonstrated that there was an interaction with the plasma membrane by performing a subcellular fractionation with differential detergent fractionation (Fig. 4e) and observed enrichment of mCherry:BHF in the membrane fraction while mCherry was only observed in the cytoplasm.

BHF resides on the intracellular side of the plasma membrane

We next investigated if recombinant BHF was localized to the intracellular or extracellular side of the plasma membrane in cycloheximide treated cells (Fig. 5). We utilized an antibody against mCherry or GPR-50 to stain the cells in the absence (Fig. 5a–d) or presence (Fig. 5a'-d') of Triton X-100 which permeabilizes the plasma membrane. HEK293T (Fig. 5a, a') and cytosolic mCherry expressing cells (Fig. 5b, b') were stained with an α-GPR-50 or α-mCherry antibody respectively as controls for the immunostaining procedure. The α-GPR-50 antibody will recognize its endogenous epitope located on the intracellular side of the plasma membrane. In the absence of Triton X-100 (Fig. 5a), puncta can be observed along the cell's border indicating that the membrane was partially permeabilized compared to Triton X-100 treated cells (Fig. 5a'). We also compared N-terminal mCherry:BHF (Fig. 5c, c') and a C-terminal fused BHF (BHF:mCherry) expressing cells (Fig. 5d, d') were used to assess if either one of the termini was presented extracellularly. Significant colocalization between the antibody and mCherry (yellow) was only observed when the cells were permeabilized with Triton X-100 (Fig. 5c', d'), indicating that the BHF localized on the intracellular side of the plasma membrane. However, we did notice some colocalization in regions of fixed cells (Fig. 5c), most likely due to formaldehyde permeabilizing part of the plasma membrane as was seen in our GPR-50 control experiment (Fig. 5a, a'). We assessed localization in another experiment through a Proteinase K protection assay on live cells expressing mCherry, mCherry:BHF, or BHF:mCherry in the absence or presence NP-40 (Fig 5e). Digestion of BHF was only observed in the presence of Proteinase K and NP-40 for both mCherry:BHF and BHF:mCherry expressing cell lines; thus corroborating the immunostaining results. Under the same conditions, mCherry expressing cells produced a smaller form as indicated by the mobility shift in the Western blot. Collectively these data indicate that BHF resides on the intracellular side of the plasma membrane when expressed in HEK293T cells.

BHF truncation mutants showed plasma membrane interaction is within the N-terminus

To further characterize BHF, several different truncation mutants were generated to help address if the interaction with the plasma membrane is based on electrostatic or a post-translational modification (PTM) to the plasma membrane or a protein-protein interaction with an integral membrane protein. We arbitrarily truncated BHF either 0, 52, 102, 152 or 202 aa from the C-terminus fused to a N-terminal mCherry (Fig. 6a–e respectively)

or 0, 50, 100, 150, or 200 aa from the N-terminus fused to a C-terminal mCherry (Fig. 6f–j respectively). In Fig. 6a–e, localization of mCherry:BHF to the plasma membrane persisted up to the shortest form (mCherry:BHF CΔ202; Fig. 6e) where localization in the cytoplasm can be observed. However, when the mCherry tagged at the C-terminal side in Fig. 6f–j, nuclear localization was observed in addition to the plasma membrane localization described above. Only BHF:mCherry retained localization to the plasma membrane with a diffuse nuclear signal (Fig. 6f). Once the first 50 aa of the N-terminus was deleted (BHF NΔ50:mCherry), stronger fluorescent signal was seen in the nucleus with faint cytoplasmic staining (Fig. 6g–j). Localization of BHF:mCherry truncation mutants differed as more of the N-terminus was deleted: full length BHF:mCherry displayed a diffused staining within the nucleus (Fig. 6f); whereas in comparison, the shorter mutants appeared punctate (Fig. 6h–j).

The truncation data suggested the likelihood of a membrane anchoring site or protein interaction site within the first 50 aa of BHF. To address the possibility of a membrane anchoring site, we subjected the primary sequence to several on-line prediction websites: No prenylation site via PrePs (Maurer-Stroh and Eisenhaber 2005), nor myristoylation via Myristoylator (Bologna et al. 2004), nor GPI anchor via PredGPI (Pierleoni et al. 2008) were predicted. However, through CSS-Palm 4.0 (Ren et al. 2008), 3 possible palmitoylation sites were predicted within the first 50 aa region: Cys20, Cys33, and Cys40. We assessed if membrane association was mediated through palmitoylation by using 2-bromopalmitate, a broad inhibitor of palmitoylation (reviewed in Draper and Smith 2009), to change its localization from the plasma membrane. We pretreated cells 30 minutes with 100 μM of 2-bromopalmitate before transfecting the mCherry:BHF construct. After 24 hours post transfection and treatment of 2-bromopalmitate, we did not observe any difference in its localization (data not shown). We followed-up with a mCherry:BHF construct with the three cysteines (Cys20, Cys33, and Cys40) mutated into serines (Fig. 6k) to determine if those cysteines have a direct role in plasma membrane localization and observed a significant amount of nuclear localization compared to the control (Fig. 6l). We quantified the percentage of cells showing nuclear localization (Fig. 6m) and observed ~68% (n = 431) of cells with staining in the nucleus compared to the control with ~2% nuclear staining (n = 478). Although plasma membrane staining still persisted in this mutant construct, the cysteines in the first 50 aa have a significant role on its localization or potential structure.

BHF can also localize to the nucleus

Since the nuclear localization was unexpected (Fig. 6f–j), the sequence of BHF (isoform 1, GenBank AGS14996.1) was submitted to cNLS Mapper (Kosugi et al. 2009) and a strong predicted bipartite nuclear localization signal (NLS) was identified (197-EYLKHQWKGQGAKKARKRIR-216). This putative NLS site may be exposed because of the mCherry fluorescent protein (~29 kDa) when fused to the C-terminal end of BHF whereas the N-terminal fusion protein displayed no significant staining in the nucleus. The unexpected localization of BHF lead us to further investigate its nuclear localization by addressing the following: (1) where in the nucleus is it located? and (2) is the putative NLS indeed present and if so masked? First, we used an antibody raised against Lamin A/C to stain the nuclear lamina (Fig. 7a–b). We immunostained mCherry:BHF

expressing cells which mainly localize to the plasma membrane (Fig. 7a) as a negative control. BHF:mCherry expressing cells displayed colocalization in regions of the nuclear lamina (Fig. 7b) and internal to the Lamin A/C stain; indicating that any nuclear signal was most likely localized to the nucleoplasm. Additionally, the truncation mutants were punctated within the nucleus and perhaps could be localized to the nucleoli (Fig. 6h–j). However, when we immunostained with an antibody against Nucleolin, we did not observed consistent colocalization (data not shown). These puncta could be in nuclear bodies devoid of Nucleolin or aggregates in protein degradation.

To determine if the nuclear localization of BHF was only possible through the mCherry tag unmasking the putative NLS region, we generated two additional constructs: BHF (isoform 2):mCherry and BHF (isoform 1):FLAG (Fig. 7c–d). Isoform 2 is predicted to lack the putative NLS based on cNLS Mapper (Kosugi et al. 2009) since two key arginines are missing (Arg214 & Arg216). When BHF isoform 2:mCherry was expressed in cells, no noticeable nuclear localization was observed: only the plasma membrane localization (Fig. 7c). This suggested that the putative NLS was abolished as predicted. When we substituted mCherry with a much smaller FLAG tag to C-terminal end of isoform 1 (BHF:FLAG), localization was only along the plasma membrane (Fig. 6d). This implied that the putative NLS was unmasked only when a mCherry tag was C-terminally fused.

Discussion

We report an initial molecular evolutionary characterization of a candidate allorecognition protein in *Botryllus schlosseri*: *Botryllus histocompatibility factor* (*bhf*). First, the degree of polymorphism and evidence for selection within the Santa Barbara harbor was examined using several statistical tests. Compared to other allorecognition genes near the *bhf* locus, *bhf* was significantly less polymorphic and was unlikely to be under any balancing or direction selection. Second, a new isoform was detected based on mRNA-Seq and an assembly of ESTs. This isoform contained 213 aa identical to the originally reported form and lacks a putative NLS signal motif. Third, the expression pattern of *bhf* isoform 1 as determined by whole-mount FISH differed slightly from what was previously reported; only in juveniles (<3 months) did we observe similar expression patterns. Expression of both *bhf* and $cfuhc^{sec}$ was restricted to a subset of blood cells within the ampullae in adult whereas juveniles showed expression in the epithelium of the ampullae in addition to blood cells. Fourth, BHF was localized to the cytoplasmic side of the plasma membrane in HEK293T cells. Through the expression of truncated forms and cysteines mutant, we concluded that the first 50 amino acids probably contain a PTM or a protein binding site which was anchoring it to the plasma membrane that is cysteine mediated, but not through palmitoylation. We also found that BHF can be localized to the nucleus when mCherry was C-terminally fused, but a smaller tag (e.g., FLAG) did not promote nuclear localization. This revealed a putative NLS region that may be normally masked until exposed by either a protein or PTM induced a conformational change.

Utilizing samples from the Santa Barbara harbor population, we determined the amount of polymorphism and diversity of BHF's amino acid diversity was not as high as $cFuhc^{sec}$ or $cFuhc^{tm}$ (Fig. 1). The majority of amino acid polymorphisms observed ranged from

switching charges (e.g., Glu54Lys) to losing a charge (e.g., Glu94Gly) (Fig. 1b). However, how these amino acid polymorphisms affect the function or structure of BHF remains unanswered. Interestingly, a mono- and bipartite NLS region was predicted by cNLS Mapper (Kosugi et al. 2009) from Glu32 to Glu54, where 3 of the 15 amino acid polymorphisms reside (sites 39, 47, and 54). Depending on the polymorphisms present, the predicted NLS (mono- and bipartite) regions were weakened or abolished. Yet, at least based on the truncation mutants from the C-terminal side, there was no noticeable nuclear localization with just the first 50 amino acid fused to mCherry. Additionally, the majority of amino acid diversity resides within a lysine rich region where 15 lysines from site 49 to 97 (~31%) are concentrated. Of which, 3 of the 15 lysines are polymorphic (Lys59Glu, Lys73Asn, and Lys81Thr) but the majority of the sequences within the Santa Barbara harbor coded for lysines. We can speculate that this lysine rich region may have a function as described in other studies: (1) interaction with ubiquitin in nucleosome if in a DNA binding domain (Oh et al. 2010); (2) involvement in signaling like in Fas signaling (Rossin et al. 2010); (3) binding to DNA in the same fashion as a linker histone via an electrostatic interaction (Caterino et al. 2011); (4) interaction with ribosomes similar to Snl1 (Verghese and Morano 2012); (5) binding to phosphoinositides, which reside on the intracellular side of the plasma membrane (reviewed in Di Paolo and De Camilli 2006), through the same mechanism as a C2 domain of rabphilin 3A (Guillén et al. 2013).

In addition to the amino acid polymorphisms, there was no evidence of selection on *bhf* based on Tajima's D and Fu and Li's D* and F* statistics (Table 3) where values of all statistical variables were not significantly different from zero. This indicated that the sample population of *bhf* was not subjected to either balancing or directional selection. This is in contrast to other allorecognition systems previously studied (reviewed in Nydam and De Tomaso 2011) where all allorecognition loci, so far, have a high level of polymorphism and experience either directional or balancing selection, in all populations studied.

Furthermore, the methodology used to identify *bhf* as an allodeterminant raises concerns (reviewed in Taketa and De Tomaso 2015). Briefly, the entire *fuhc* region, encoding the *fester* family genes, *bhf*, *cfuhc*^{sec}, *cfuhc*^{tm}, and *hsp40-l* are located in a ca. 1 cM region with no recombination in 5 independent crosses; and the 1 cM *fuhc* was physically mapped to a ca. 1 Mb region (De Tomaso and Weissman 2003A; De Tomaso et al. 2005). The distance between *bhf* and *hsp40-l* is 10 times smaller; and these genes are encoded in a ca. 120 Kb region within the chromosomal breakpoints. Despite this tight linkage, Voskoboynik *et al.* (2013B) reported differences in predictability between *bhf* and *cfuhc*^{sec}, which are separated by only 62 Kb, as well as between *cfuhc*^{sec} and *cfuhc*^{tm}, which have an intergenic distance of only 247 bp. Some of their reported mismatches included pairings used in defined crosses in the same studies, but we have previously published that some individuals were incorrectly genotyped, as identified using a neutral segregating genetic marker (Nydam et al. 2013). However, they also claimed equivalent changes in wild-type pairings. It appears that this earlier genetic contamination affected their *in silico* analysis since one of their metrics, homozygous error, was utilized to calculate a classification error for a given gene. This compounding error resulted in the conclusion that *hsp40-l*, which is encoded only 8 Kb from *cfuhc*^{tm}, possesses no predictive ability (Voskoboynik, et al., 2013B). In other words, according to their *in silico* analysis, *hsp40-l* was segregating independently of genes encoded

only 8 Kb away. *A priori*, these results should have raised red flags. The correlation of genetic to physical distance from defined crosses (ca. 1Mb/cM; De Tomaso and Weissman 2003A) predicts that the probability of a recombination event between *bhf* and *cfuhc* is conservatively 1/1000 per generation, and between *cfuhc*^{sec} and *cfuhc*^{tm} 1/400,000. More importantly, the complete lack of linkage disequilibrium of any of these genes with *hsp40-1* is highly unlikely. Thus the *in silico* analysis identifying multiple mismatches from 29 pairings is unlikely to be correct, and the role of *bhf in* allorecognition needs to be further demonstrated at a biochemical level.

Localization of *bhf* through whole-mount FISH utilizing two probes revealed a small population of cells within the ampullae and other tissues (Fig. 3a–d) in adult colonies. This differed from previous observations in Voskoboynik *et al*. (2013B) where they reported expression of *bhf* within the blood vessels and epithelium of the ampullae. However, through double labeled whole-mount FISH with *bhf* and *cfuhc*^{sec}, we observed expression patterns for *bhf as* previously reported (Voskoboynik et al. 2013B) but only in juvenile colonies (<3 months; Fig. 6e–h’’’); whereas adult colonies were comparable to the single labeled whole-mount FISH (data not shown). We also attempted a double label FISH of *bhf* in conjunction with *fester* or *uncle fester*; however, no significant signal was observed for either *fester* or *uncle fester* (data not shown). Additionally, the expression pattern for *cfuhc*^{sec} differed from our previous report (Nydam et al. 2013); we did not observe any significant signal in epithelial cells of the ampullae (Fig. 3e’). These results are most likely due to the limitations of our new FISH protocol (Langenbacher et al. 2015) which has greatly improved localization and reduced background, but is significantly less sensitive. For example, Fester and Uncle Fester protein is detected at the tips of the ampullae using immunofluorescence, along with the corresponding mRNA detected by other *in situ* methods (Nyholm et al. 2006; McKitrick et al. 2011; summarized in Online Resource 1, Fig. S1). However, these genes are also expressed at significantly lower levels than *bhf*. In addition, the age, genotype, and blastogenic stage of the colony may contribute to these differences considering we observed an expression pattern difference between adult and juvenile colonies. This also suggests that *bhf* could also be a temporal and/or genotype-specific expressed gene. This would require another study utilizing several well-defined genetic lines and their off-springs to address this possibility.

Interestingly in the double labeled FISH, there was a subset of blood cells which expressed both *bhf* and *cfuhc*^{sec} (white arrows; Fig. 3e–g’’’). We observed these double labeled cells within the ampullae (Fig. 3e’’’), blood vessel (Fig. 3f’’), and around the endostyle (Fig. 3g’’’). However, there may be co-expressing cells along the epithelium of the ampullae are not detected with our protocol. Nonetheless, the function and role of these double labeled cells remains uncertain. Also, determining the exact lineage of these blood cells still remains a technical challenge, but we are in the process of identifying and development robust lineage-specific *in situ* markers. Given the current whole-mount FISH (Fig. 3), we speculate that a population of these *bhf* expressing blood cells may be macrophage-like cells given that there were *bhf* expressing cells observed in resorbing zooids (Fig. 3b) where macrophage-like and blood cells are abundant (Lauzon et al. 1992). Macrophage-like cells are capable of migrating through the ampullae epithelium (Rinkevich et al. 1998), which may explain the identity of the *bhf* expressing cells along the epithelium of the ampullae

(Fig. 3a, d) in adult colonies. Additionally, the frequency of circulating macrophage-like cells was previously reported to increase during takeover (stage D) when measured by cytofluorimetric analysis (Ballarin et al. 2008). This is also consistent with the mRNA-Seq analysis where there was an increase of expression in stage D of the blastogenic cycle (Fig. 2c, d) in adult colonies.

Ectopic expression of truncated mutant fusion proteins in cell culture systems has been used in previous studies to identify and characterize putative domains within a protein (Moede et al. 1999; Zhang et al. 2002; Sheng et al. 2004). Using a similar approach in HEK293T, several putative motifs were identified (NLS, possible PTM via cysteines, and a lysine-rich region) that localized BHF to the cytoplasmic side of the plasma membrane (Fig. 5) and nucleus (Fig. 6f). However, we are still investigating how BHF is associating with the plasma membrane. We initially thought that BHF was anchored to the plasma membrane through palmitoylation of one or more cysteines within the first 50 aa region based on the bioinformatics results from several databases, but drug treatment with 2-bromopalmitate resulted in no difference in localization (data not shown) while the cysteine mutant produced a nuclear localization but did not abolish the plasma membrane localization (Fig. 5k). Additionally, the cysteines in that region are most likely reduced given the redox potential in the cytoplasm (reviewed in López-Mirabel and Winther 2008). Thus, the cysteines were likely not involved in structural disulfide bridges unless trafficked through the rough ER. If so, BHF has to be trafficked in a non-canonical method since it lacks a predicted signal peptide. The possibility of the lysine rich region (within aa position 50–100) interacting with the plasma membrane either through electrostatic interaction or via a protein-protein interaction is unlikely given that mCherry:BHF CΔ202 still resides at the plasma membrane (Fig. 6e) and complete nuclear localization was observed when the first 50 aa was removed (BHF NΔ50:mCherry; Fig. 6g). Additionally, BHF was mainly detected on the intracellular side of the plasma membrane through immunostaining (Fig. 5c–d') and Proteinase K protection assay (Fig. 5e). This suggested that its role in allorecognition as an allodeterminant was less likely, but instead might be involved in the downstream signaling due to its putative NLS region or acting as a scaffold at the plasma membrane for other allorecognition players like Fester (Nyholm et al. 2006) or Uncle Fester (McKitrick et al. 2011). However, BHF's role as a possible surface recognition marker cannot be eliminated since there could be a specific *Botryllus* protein that will allow it to be displayed on the extracellular side.

Interestingly, when cells expressing BHF:mCherry were treated with cycloheximide, an inhibitor of translation, localization was only observed at the plasma membrane (Fig. 5d, d'). However, when the same cells were not treated, localization was observed at the plasma membrane and nucleus because a putative NLS sequence was exposed when fused to a C-terminal mCherry (Fig. 6f, Fig. 7b). Most likely, due to expressing BHF:mCherry ectopically, there was saturation of the machinery or protein anchoring it to the plasma membrane thereby allowing the nuclear import machinery to transport BHF into the nucleus. Within the nucleus, the half-life of BHF:mCherry might be heavily reduced compared to the plasma membrane population and could be degraded via ubiquitination. A few sites have been strongly predicted to be ubiquitinated by UbPred (Radivojac et al. 2010): Lys22, Lys59, and Lys187. Since BHF has been previously analyzed to contain disordered regions

(Voskoboynik et al. 2013B), perhaps degradation within the nucleus might be analogous to the yeast's protein quality control degradation system (Rosenbaum et al. 2011) where San1 homolog recognizes disordered proteins within the nucleus and ubiquitinates them. This putative NLS sequence was further confirmed when we identified (Fig. 2a, b) and expressed a shorter, second isoform of BHF (Fig. 4b, Fig. 7c). This 219 aa isoform lacked 2 arginine residues thereby disrupting the NLS sequence. When expressed with a C-terminal mCherry, no significant localization was observed in the nucleus (Fig. 7c). Additionally, the putative NLS was cryptic since nuclear localization was observed when fused with mCherry (Fig. 7b). When the mCherry tag was replaced with a smaller FLAG tag, no significant nuclear localization was detected (Fig. 7d).

Without the biochemical data from *B. schlosseri*, we cannot conclusively determine if BHF is an allodeterminant. However, if BHF is not acting as an allodeterminant and given the data thus far, we speculate that BHF could function as a scaffolding protein in a complex with other *Botryllus* proteins. If this complex includes an allorecognition protein (e.g. Fester and/or Uncle Fester), then knockdown of this complex would be consistent with the 'no reaction' phenotype reported (Voskoboynik et al. 2013B). BHF could also play a role in signal transduction due to its C-terminal NLS and allow it to translocate into the nucleus. Within the nucleus, it may function either in chromatin remodeling or gene regulation as seen in other examples with a lysine-rich region (discussed above). Meanwhile, the shorter isoform may exist to modulate the amount of BHF (isoform 1) translocating into the nucleus by acting simply as a scaffold at the plasma membrane.

Further identification and characterization of the allodeterminants in *Botryllus schlosseri* will be pivotal for understanding how allorecognition functions in invertebrates. The discovery of *bhf* (Voskoboynik et al. 2013B) and the data supporting its involvement in *B. schlosseri's* allorecognition made it an interesting candidate for further molecular and cellular characterization. Future studies will investigate if there are any interactions with the known allorecognition players; whether BHF binds to DNA and regulates gene expression; and will determine its binding partners and how it associates with the plasma membrane.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Michael B. Caun for his expert care of the De Tomaso laboratory mariculture facility. We also thank Dr. Kathy Foltz and Dr. Jeffrey Bailey for their critical comments on this manuscript. They also thank Liviu Cengher and Michael A. Trebino for their assistance generating several constructs. They acknowledge the use of the NRI-MCDB Microscopy Facility and the Spectral Laser Scanning Confocal supported by the Office of The Director, National Institutes of Health of the NIH under Award # 1 S10 OD010610-01A1. This work was funded by the following grants through the National Institutes of Health: AI041588 (to AWD) and F32 GM108227 (to ADL). This work was also funded by California Institute for Regenerative Medicine: T3-00009 (to DR).

References

Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11: R106. [PubMed: 20979621]

Ballarin L, Cima F. Cytochemical properties of Botryllus schlosseri haemocytes: indications for morpho-functional characterisation. Eur J Histochem. 2005; 49 (3) 255–264. [PubMed: 16216811]

Ballarin L, Menin A, Tallandini L, Matozzo V, Burighel P, Basso G, Fortunato E, Cima F. Haemocytes and blastogenetic cycle in the colonial ascidian Botryllus schlosseri: a matter of life and death. Cell Tissue Res. 2008; 331: 555–564. [PubMed: 17972103]

Boehm T, McCurley N, Sutoh Y, Schorpp M, Kasahara M, Cooper MD. VLR-based adaptive immunity. Annu Rev Immunol. 2012; 30: 203–220. [PubMed: 22224775]

Bologna G, Yvon C, Duvaud S, Veuthey AL. N-terminal myristoylation predictions by ensembles of neural networks. Proteomics. 2004; 4: 1626–1632. [PubMed: 15174132]

Braden BP, Taketa DA, Pierce JD, Kassmer S, Lewis DD, De Tomaso AW. Vascular regeneration in a basal chordate is due to the presence of immobile, bi-functional cells. PLoS ONE. 2014; 9 (4) e95460. [PubMed: 24736432]

Burnet FM. “Self-recognition” in colonial marine forms and flowering plants in relation to the evolution of immunity. Nature. 1971; 232: 230–235. [PubMed: 4937075]

Caterino TL, Fang H, Hayes JJ. Nucleosome linker DNA contacts and induces specific folding of the intrinsically disordered H1 carboxyl-terminal domain. Mol Cell Biol. 2011; 31 (11) 2341–2348. [PubMed: 21464206]

Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature. 2006; 439 (7079) 965–968. [PubMed: 16495997]

Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H. Additional molecular support for the new chordate phylogeny. Genesis. 2008; 46 (11) 592–604. [PubMed: 19003928]

De Tomaso AW, Weissman IL. Initial characterization of a protochordate histocompatibility locus. Immunogenetics. 2003A; 55: 480–490. [PubMed: 14520503]

De Tomaso AW, Weissman IL. Construction and characterization of large-insert genomic libraries (BAC and fosmid) from the ascidian Botryllus schlosseri and initial physical mapping of a histocompatibility locus. Mar Biotechnol. 2003B; 5: 103–115.

De Tomaso AW, Nyholm SV, Palmeri KJ, Ishizuka KJ, Ludington WB, Mitchel K, Weissman IL. Isolation and characterization of a protochordate histocompatibility locus. Nature. 2005; 438 (7067) 454–459. [PubMed: 16306984]

De Tomaso AW. Allorecognition, germline chimerism, and stem cell parasitism in the colonial ascidian, Botryllus schlosseri. Biol Theory. 2014.

Di Paolo G, De Camilli P. Phosphoinositides in cell regulation and membrane dynamics. Nature. 2006; 443: 651–657. [PubMed: 17035995]

Draper JM, Smith CD. Palmitoyl acyltransferase assays and inhibitors (Review). Molec Membrane Biol. 2009; 26 (1) 5–13.

Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics. 1993; 133: 693–709. [PubMed: 8454210]

Grosberg RK. The evolution of allorecognition specificity in clonal invertebrates. Q Rev Biol. 1988; 63 (4) 377–412.

Guillén J, Ferrer-Orta C, Buxaderas M, Pérez-Sánchez D, Guerrero-Valero M, Luengo-Gil G, Pous J, Guerra P, Gómez-Fernández JC, Verdaguer N, Corbalán-García S. Structural insights into Ca^{2+} and $PI(4,5)P_2$ binding modes of the C2 domains of rabphilin 3A and synaptotagmin 1. Proc Natl Acad Sci USA. 2013; 110 (51) 20503–20508. [PubMed: 24302762]

Heo WD, Inoue T, Park WS, Kim ML, Park BO, Wandless TJ, Meyer T. $PI(3,4,5)P_3$ and $PI(4,5)P_2$ lipids target proteins with polybasic clusters to the plasma membrane. Science. 2006; 314 (5804) 1458–1461. [PubMed: 17095657]

Kawai N, Shimada M, Kawahara H, Satoh N, Yokosawa H. Regulation of ascidian Rel by its alternative splice variant. Eur. J. Biochem. 2003; 270 (22) 4459–4468. [PubMed: 14622274]

Kondoh M, Kasai T, Shimada M, Kashiwayanagi M, Yokosawa H. cDNA cloning and characterization of an osmotically sensitive TRP channel from ascidian eggs. Comp Biochem Physiol B, Biochem Mol Biol. 2003; 134: 417–423. [PubMed: 12628373]

Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of yeast cell cycle-dependent nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc Natl Acad Sci USA. 2009; 106: 10171–10176. [PubMed: 19520826]

Laird DJ, De Tomaso AW, Weissman IL. Stem cells are units of natural selection in a colonial ascidian. Cell. 2005; 123: 1351–1360. [PubMed: 16377573]

Langenbacher AD, Rodriguez D, Di Maio A, De Tomaso AW. Whole-mount fluorescent in situ hybridization staining of the colonial tunicate Botryllus schlosseri. Genesis. 2015; 53 (1) 194–201. [PubMed: 25179474]

Lauzon RJ, Ishizuka KJ, Weissman IL. A cyclical, developmentally-regulated death phenomenon in a colonial urochordate. Dev Dynam. 1992; 194: 71–83.

Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25: 1451–1452. [PubMed: 19346325]

Litman GW, Dishaw LJ. Histocompatibility: Clarifying fusion confusion. Current Biology. 2013; 23 (20) R934–R935. [PubMed: 24156816]

Long EO, Kim HS, Liu D, Peterson ME, Rajagopalan S. Controlling NK cell responses: integration of signals for activation and inhibition. Annu Rev Immunol. 2013; 31: 227–258. [PubMed: 23516982]

López-Mirabel HR, Winther JR. Redox characteristics of the eukaryotic cytosol. Biochem Biophys Acta. 2008; 1783 (4) 629–640. [PubMed: 18039473]

Maurer-Stroh S, Eisenhaber F. Refinement and prediction of protein prenylation motifs. Genome Biol. 2005; 6 (6) R55. [PubMed: 15960807]

McKitrick TR, Muscat CC, Pierce JD, Bhattacharya D, De Tomaso AW. Allorecognition in a basal chordate consists of independent activating and inhibitory pathways. Immunity. 2011; 34: 616–626. [PubMed: 21497115]

Zhang T, Leibiger B, Pour HG, Berggren P, Leibiger IB. Identification of a nuclear localization signal, RRMKWKK, in the homeodomain transcription factor PDX-1. FEBS Letters. 1999; 461: 229–234. [PubMed: 10567702]

Nydam ML, De Tomaso AW. Creation and maintenance of variation in allorecognition loci: molecular analysis in various model systems. Front Immunol. 2011; 2: 79. [PubMed: 22566868]

Nydam ML, Taylor AA, De Tomaso AW. Evidence for selection on a chordate histocompatibility locus. Evolution. 2012; 67 (2) 487–500. [PubMed: 23356620]

Nydam ML, Netuschil N, Sanders E, Langenbacher A, Lewis DD, Taketa DA, Marimuthu A, Gracey AY, De Tomaso AW. The candidate histocompatibility locus of a basal chordate encodes two highly polymorphic proteins. PLoS ONE. 2013; 8 (6) e65980. [PubMed: 23826085]

Nyholm SV, Passegue E, Ludington WB, Voskoboynik A, Mitchel K, Weissman IL, De Tomaso AW. fester, a candidate allorecognition receptor from a primitive chordate. Immunity. 2006; 25: 163–173. [PubMed: 16860765]

Oh S, Jeong K, Kim H, Kwon CS, Lee D. A lysine-rich region in Dot1p is crucial for direct interaction with H2B ubiquitylation and high level methylation of H3K79. Biochem Biophys Res Commun. 2010; 339: 512–517.

Pierleoni A, Martelli PL, Casadio R. PredGPI: a GPI-anchor predictor. BMC bioinformatics. 2008; 9: 392. [PubMed: 18811934]

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Austria: Vienna; 2014. http://www.R-project.org/

Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, Goebi MG, Iakoucheva LM. Identification, analysis, and prediction of protein ubiquitination sites. Proteins. 2010; 78 (2) 365–380. [PubMed: 19722269]

Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X. CSS-Palm 2.0: an updated software for palmitoylation site prediction. Protein Eng Des Sel. 2008; 21 (11) 639–644. [PubMed: 18753194]

Rinkevich B, Tartakover S, Gershon H. Contribution of morula cells to allogenic responses in the colonial urochordate Botryllus schlosser*i*. Mar Biol. 1998; 131 (2) 227–236.

Rinkevich B, Douek J, Rabinowitz C, Paz G. The candidate Fu/HC gene in Botryllus schlosseri (Urochordata) and ascidians' historecognition – An oxymoron? Dev Comp Immunol. 2012; 36 (4) 718–727. [PubMed: 22085780]

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011; 29: 24–26. [PubMed: 21221095]

Rodi DJ, Mandava S, Makowski L. DIVAA: analysis of amino acid diversity in multiple aligned protein sequences. Bioinformatics. 2004; 20: 3481–3489. [PubMed: 15284106]

Rodriguez D, Sanders EN, Farell K, Langenbacher AD, Taketa DA, Hopper MR, Kennedy M, Gracey A, De Tomaso AW. Analysis of the basal chordate Botryllus schlosseri reveals a set of genes associated with fertility. BMC Genomics. 2014; 15 (1) 1183. [PubMed: 25542255]

Rosenbaum JC, Fredrickson EK, Oeser ML, Garrett-Engele CM, Locke MN, Richardson LA, Nelson ZW, Hetrick ED, Milac TI, Gottschling DE, Gardner RG. Disorder targets misorder in nuclear quality control degradation: a disordered ubiquitin ligase directly recognizes its misfolded substrates. Mol Cell. 2011; 41: 93–106. [PubMed: 21211726]

Rosengarten RD, Nicotra ML. Model systems of invertebrate allorecognition. Curr Biol. 2011; 21: R82–R92. [PubMed: 21256442]

Rossin A, Kral R, Lounnas N, Chakrabandhu K, Mailfert S, Marguet D, Hueber A. Identification of a lysine-rich region of Fas as a raft nanodomain targeting signal necessary for Fas-mediated cell death. Exp Cell Res. 2010; 316: 1513–1522. [PubMed: 20298688]

Sasaki N, Ogasawara M, Sekiguchi T, Kusumoto S, Satake H. Toll-like receptors of the ascidian *Ciona intestinalis:* Prototypes with hybrid functionalities of vertebrate Toll-like receptors. J Biol Chem. 2009; 284 (40) 27336–27343. [PubMed: 19651780]

Schlumpberger JM, Weissman IL, Scofield VL. Separation and labeling of specific subpopulations of Botryllus blood cells. J Exp Zool. 1984; 229: 401–411. [PubMed: 6368739]

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Figi: an open-source platform for biological-image analysis. Nature Methods. 2012; 9 (7) 676–682. [PubMed: 22743772]

Scofield VL, Schlumpberger JM, West LA, Weissman IL. Protochordate allorecognition is controlled by a MHC-like gene system. Nature. 1982; 295: 499–502. [PubMed: 7057909]

Sheng Z, Lewis JA, Chirico WJ. Nuclear and nucleolar localization of 18-kDa fibroblast growth factor-2 is controlled by C-terminal signals. J Biol Chem. 2004; 279 (38) 40153–40160. [PubMed: 15247275]

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123: 585–595. [PubMed: 2513255]

Taketa DA, De Tomaso AW. Botryllus schlosseri allorecognition: tackling the enigma. Dev Comp Immunol. 2015; 48 (1) 254–265. [PubMed: 24709050]

Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics. 2012; 14 (2) 178–192. [PubMed: 22517427]

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechol. 2010; 28: 5111–515.

Verghese J, Morano KA. A lysine-rich region within fungal BAG domain-containing proteins mediates a novel association with ribosomes. Eukaryotic Cell. 2012; 11 (8) 1003–1011. [PubMed: 22635919]

Voskoboynik A, Neff NF, Sahoo D, Newman AM, Pushkarev D, Koh W, Passarelli B, Fan HC, Mantalas GL, Palmeri KJ, Ishizuka KJ, Gissi C, Griggio F, Ben-Shlomo R, Corey DM, Penland L, White RA, Weissman IL, Quake SR. The genome sequence of the colonial chordate, Botryllus schlosseri. eLife. 2013A; 2: e00569. [PubMed: 23840927]

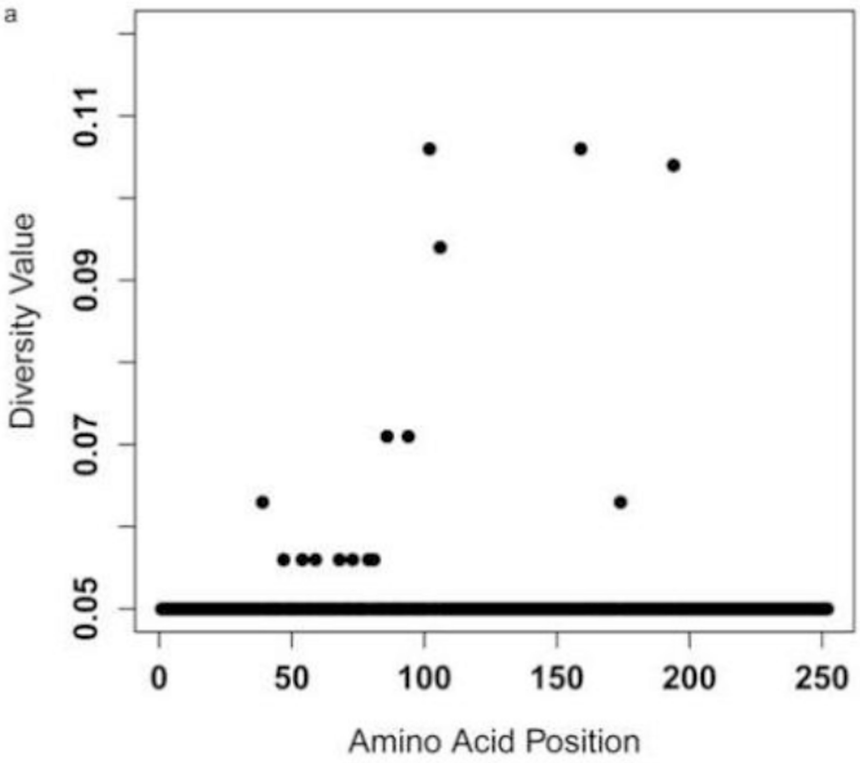
Voskoboynik A, Newman AM, Corey DM, Sahoo D, Pushkarev D, Neff NF, Passarelli B, Koh W, Ishizuka KJ, Palmeri KJ, Dimov IK, Keasar C, Fan HC, Mantalas GL, Sinha R, Penland L, Quake SR, Weissman IL. Identification of a colonial chordate histocompatibility gene. Science. 2013B; 341: 384–387. [PubMed: 23888037]

Weissman IL, Saito Y, Rinkevich B. Allorecognition histocompatibility in a protochordate species: is the relationship to MHC somatic or structural? Immunol Rev. 1990; 113: 227–241. [PubMed: 2180808]

Wilson DJ, McVean G. Estimating diversifying selection and functional constraint in the presence of recombination. Genetics. 2006; 172: 1411–1425. [PubMed: 16387887]

Zhang YA, Okada A, Lew CH, McConnell SK. Regulated nuclear trafficking of the homeodomain protein Otx1 in cortical neurons. Mol Cell Neurosci. 2002; 19: 430–446. [PubMed: 11906214]

Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol. 2000; 7 (1–2) 203–214. [PubMed: 10890397]



b

```
  1 MVHDTEQLLA QGHHEEETEC GKYGKLPEKG SECKKHGIFC RILTALHLKK 50
                                          L        R

 51 RRTEHDHQKL LSESQEHLDA STKKTKKKAK KDKRKNKPPK KDSETSKPAQ 100
      K    E          I     N     S  T    I         G

101 TTISRLPSNR NNNNANSFAT TYEKFDNDSL CSVDLIPVDI EFWDMENEPV 150
     A   F

151 DQLPHEILES VHMYGDDRFG ERLIDRAQNK YAPLDEKQRS ESHGAGEYLK 200
           Q                V                      S

201 HQWKGQGAKK ARKRIRTVMK ATWQSLQAGA RSQTAFLNPQ GAVSAALVQN 250


251 RR                                                   252
```

Fig. 1.

Amino acid diversity of BHF. (a) DIVAA analysis (Rodi et al. 2004) of BHF sequences shows amino acid diversity of individuals in the Santa Barbara harbor based on 29 sequences. (b) Sequence of BHF (GenBank AGS14996.1) with the amino acid polymorphisms found in the Santa Barbara harbor population shown below. Bolded text represents significant changes in amino acid residues' properties

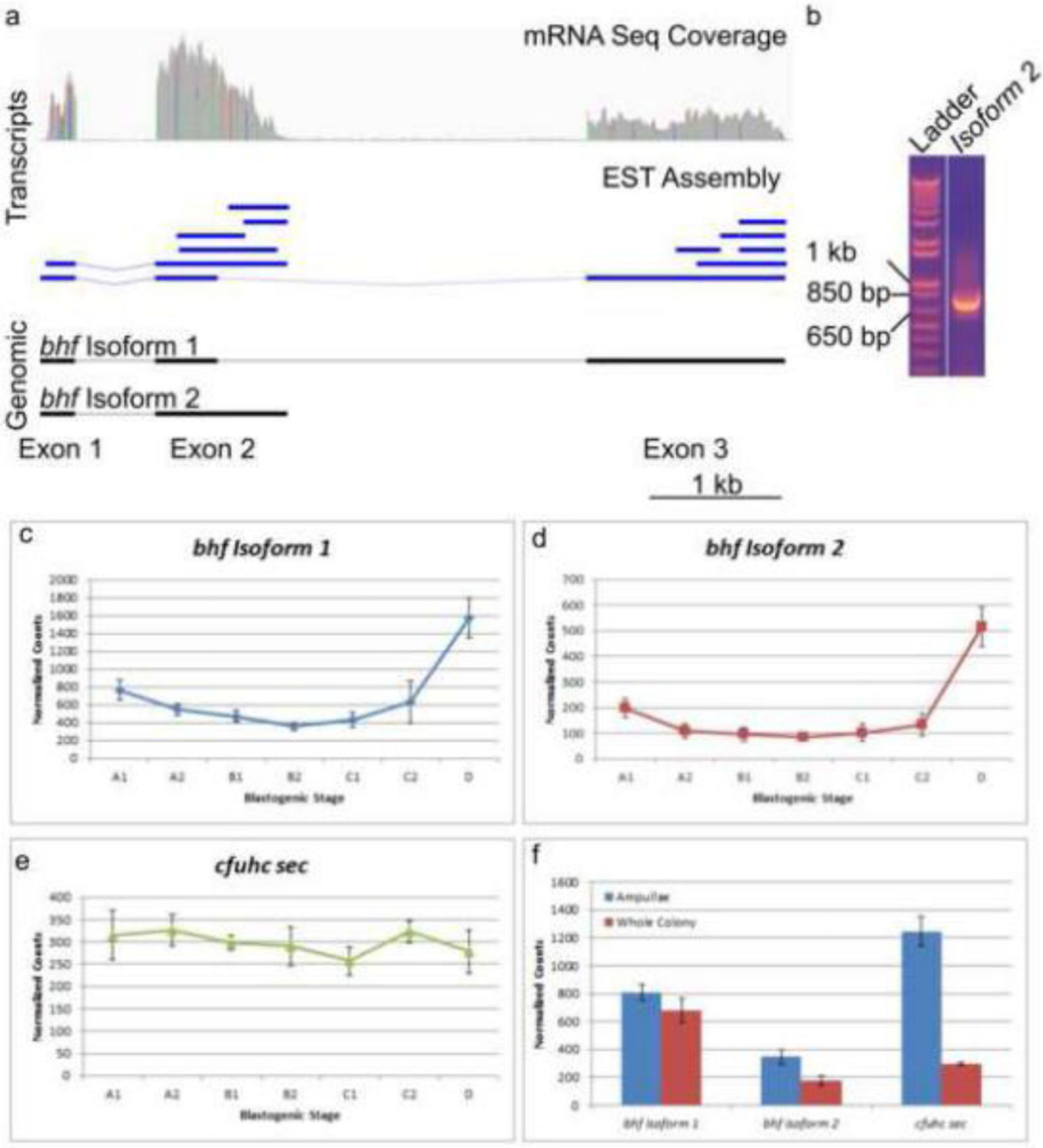


Fig. 2.

mRNA-Seq expression profile. (a) Schematic representation of genomic structure (black) with overlapping EST hits (blue) with an E-value < $1e^{-10}$ and mRNA-Seq reads mapped to the genomic region visualized on IGV. Gray region in the mRNA-Seq represent conserved nucleotide while the colors represent the following nucleotides: red (T), blue (C), green (A), and orange (G). (b) RT-PCR of *bhf* isoform 2 on a 1% agarose gel with an expected size of 718 bp. (c–e) mRNA-Seq expression profile based on normalized counts across the blastogenic cycle (n = 4 per stage) for *bhf* isoform 1 (c), isoform 2 (d), $cfuhc^{sec}$ (e). (f)

mRNA-Seq between ampullae tissue (n = 15) and whole colony (n = 28) for *bhf* isoform 1, isoform 2, and *cfuhc*^{sec}. Error bars are standard error of normalized counts from DESeq with a false discovery rate = 10%

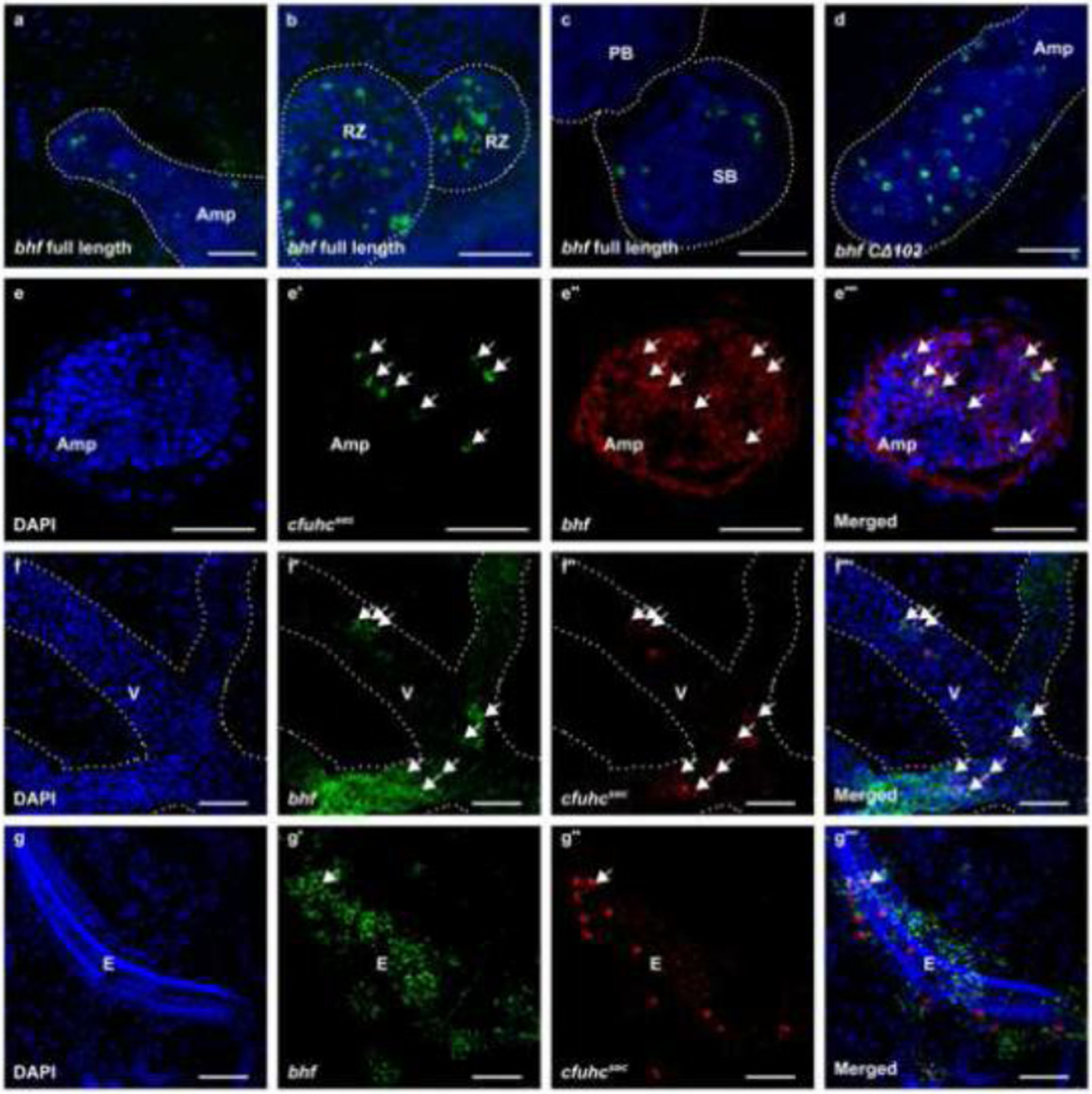


Fig. 3.

Expression localization via whole-mount FISH. (a–d) Confocal Z-stacks of adult colonies with either a *bhf* isoform 1 full length probe (a–c) or a *bhf* CΔ102 truncated probe that will bind both isoforms (d) on ampullae (a & d), resorbing zooid (blastogenic cycle stage D; b), and in the secondary bud (c). (e-g''') Confocal images on double labeled whole-mount FISH on juvenile colonies (<3 months; ~3–4 zooids) with probes designed against *bhf* CΔ102 and *cfuhc*^{sec}. (e–e''') Confocal images of an ampullae. (f–f''') Z-stack of blood vessels. (g–g''') Z-stack of the endostyle. All samples were counterstained with DAPI (blue). White arrows

indicate co-expressing cells of *bhf* CΔ102 and *cfuhc*^{sec}. Dashed outlines mark the boundary of the indicated tissue. Amp = ampullae; RZ = resorbing zooid; E = endostyle; PB = primary bud; SB = secondary bud; V = blood vessel. Scale bar = 50 μm

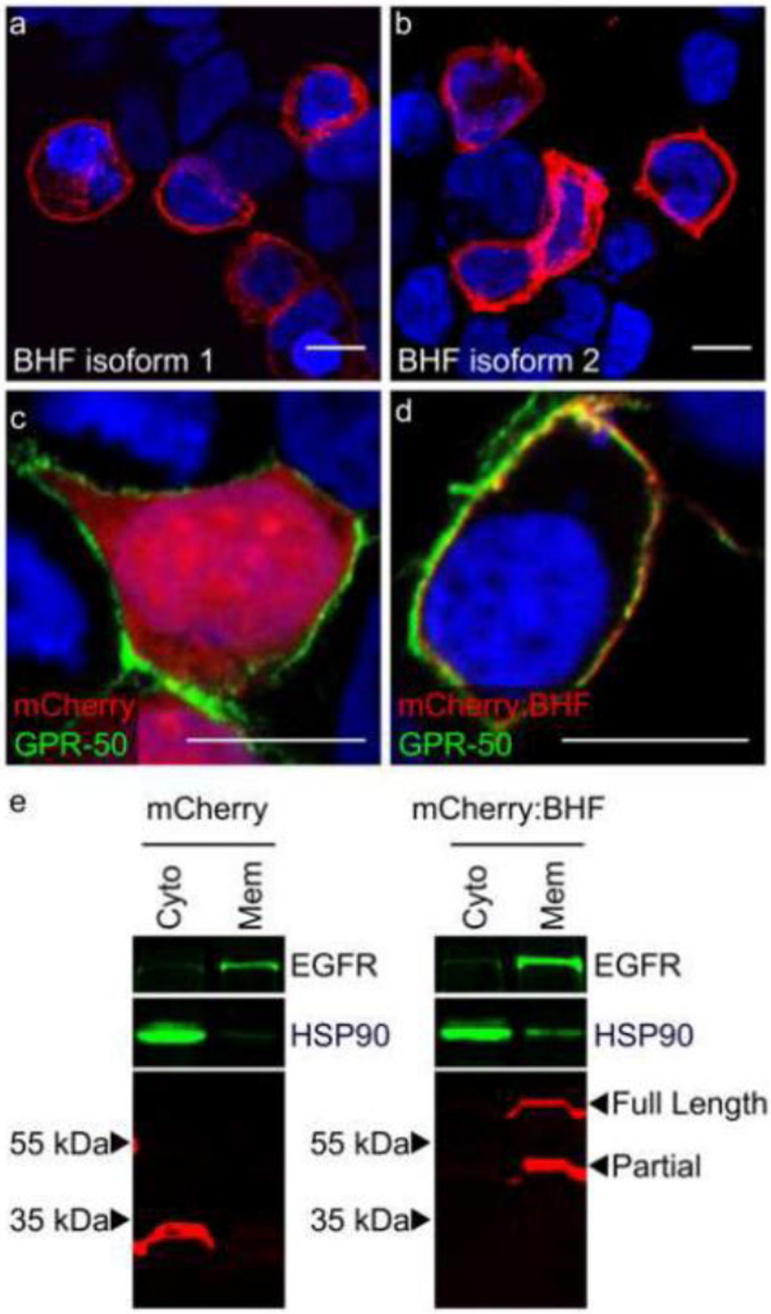


Fig. 4.

Localization of BHF in HEK293T cells. (a–b) Confocal Z-stack image of mCherry:BHF (isoform 1) (a) or mCherry:BHF (isoform 2) (b). (c–d) Confocal slices of mCherry (c) or mCherry:BHF isoform 1 (d) expressing HEK293T cells (red) stained with a rabbit α-GPR50 antibody (green). All images have been counterstained with DAPI (blue). (e) Western blot of subcellular fractions (cytosol = Cyto; membrane = Mem) on transiently expressing mCherry (~29 kDa) or mCherry:BHF isoform 1 (~56 kDa) cells blotted with a rat α-mCherry, rabbit

α-EGFR, or rabbit α-HSP90 antibody. The lower bands (~45 kDa) in the right panel are truncated mCherry:BHF from degradation or partially translated. Scale bar = 10 μm

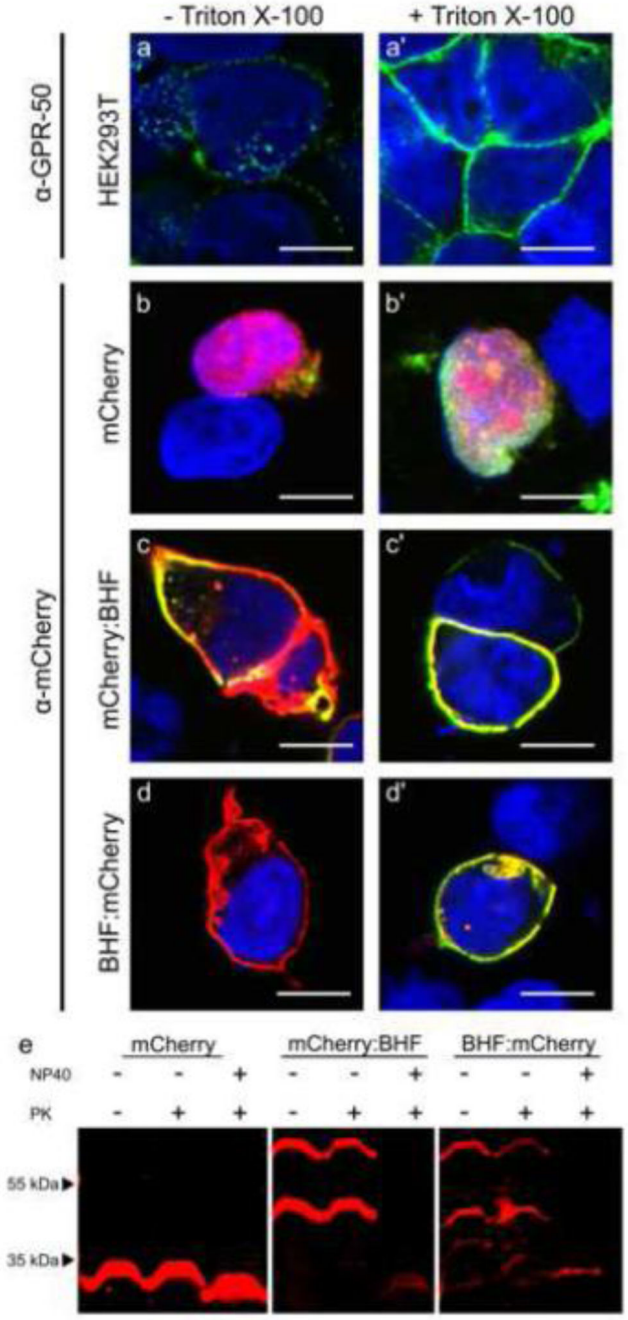


Fig. 5.

BHF is localized to the inner side of the plasma membrane in cycloheximide treated HEK293T (a–d) and proteinase K protection assay (e). (a, a') HEK293T negative control stained for endogenous GPR-50 (green) in the absence (a) or presence (a') of Triton X-100. The epitope for GPR-50 is on the intracellular side of the plasma membrane. (b, b') mCherry (red) transient expression. (c, c') mCherry:BHF (isoform 1; red) transient expression. (d, d') BHF:mCherry (isoform 1; red) transient expression. (b, c, d) Negative controls stained with a rat α-mCherry (green) and counterstained with DAPI (blue). (b', c', d') Permeabilized

with Triton X-100 then stained with a rat α-mCherry (green), mCherry (red), DAPI (blue). Scale bar = 10 μm. (e) Western blot from a proteinase K protection assay using a rat α-mCherry antibody to detect recombinant proteins in transient expressing cells

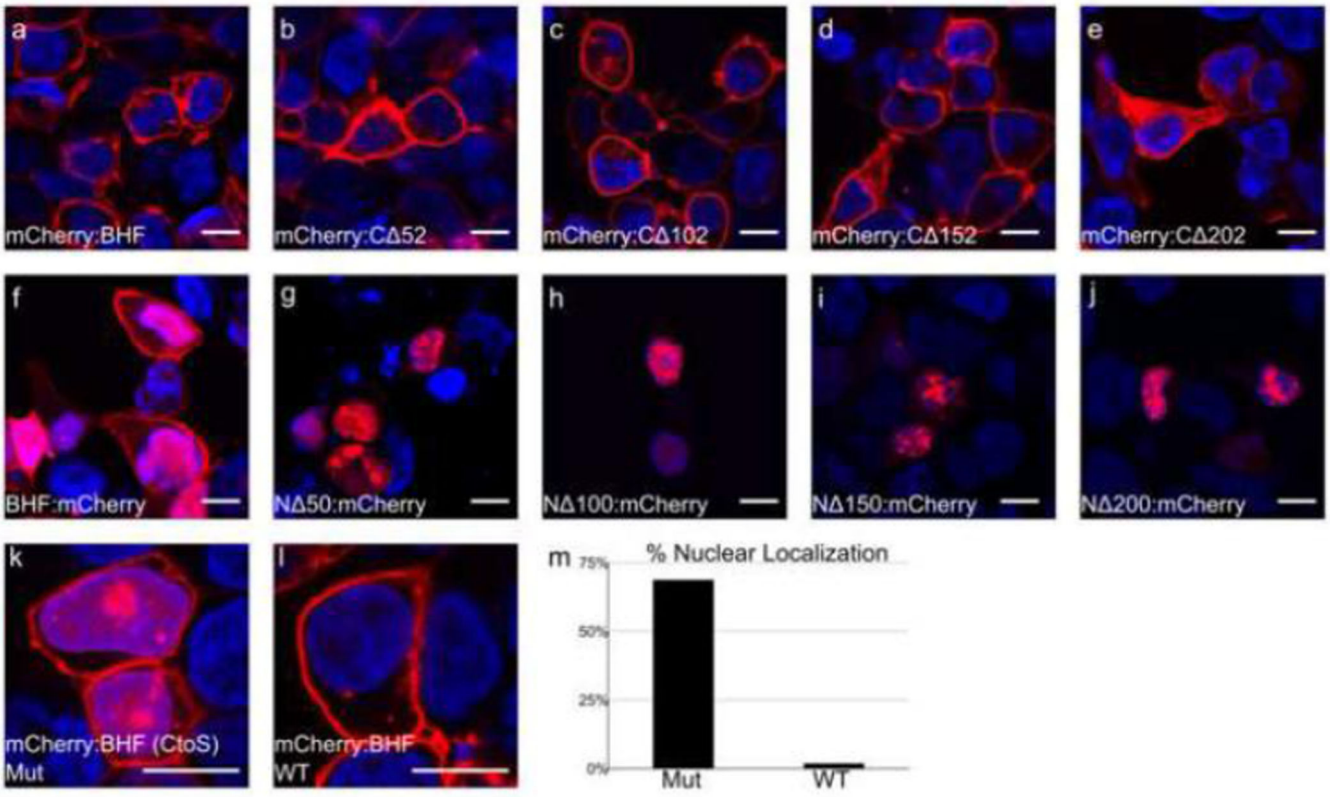


Fig. 6.

Confocal images of different truncation mutants of mCherry:BHF (a–f) or BHF:mCherry (f–j) in red. NΔ or CΔ represent which amino acid terminal end was truncated by *x* many amino acids. (a) Full length mCherry:BHF. (b) BHF CΔ52:mCherry. (c) BHF CΔ102:mCherry. (d) BHF CΔ152:mCherry. (e) BHF CΔ202:mCherry. (f) Full length BHF:mCherry. (g) NΔ50 BHF:mCherry. (h) NΔ100 BHF:mCherry. (i) NΔ150 BHF:mCherry. (j) NΔ200 BHF:mCherry. (k) Mutant construct of the 3 cysteines into serines within the first 50 aa of mCherry:BHF acquired with a 60× objective. (l) Same construct as (a) but repeated and acquired with a 60× objective. (m) Percentage of cells that showed nuclear localization in the mutant (n = 431 cells) and wildtype control (n = 478 cells). All images were counterstained with DAPI (blue). Scale bar = 10 μm

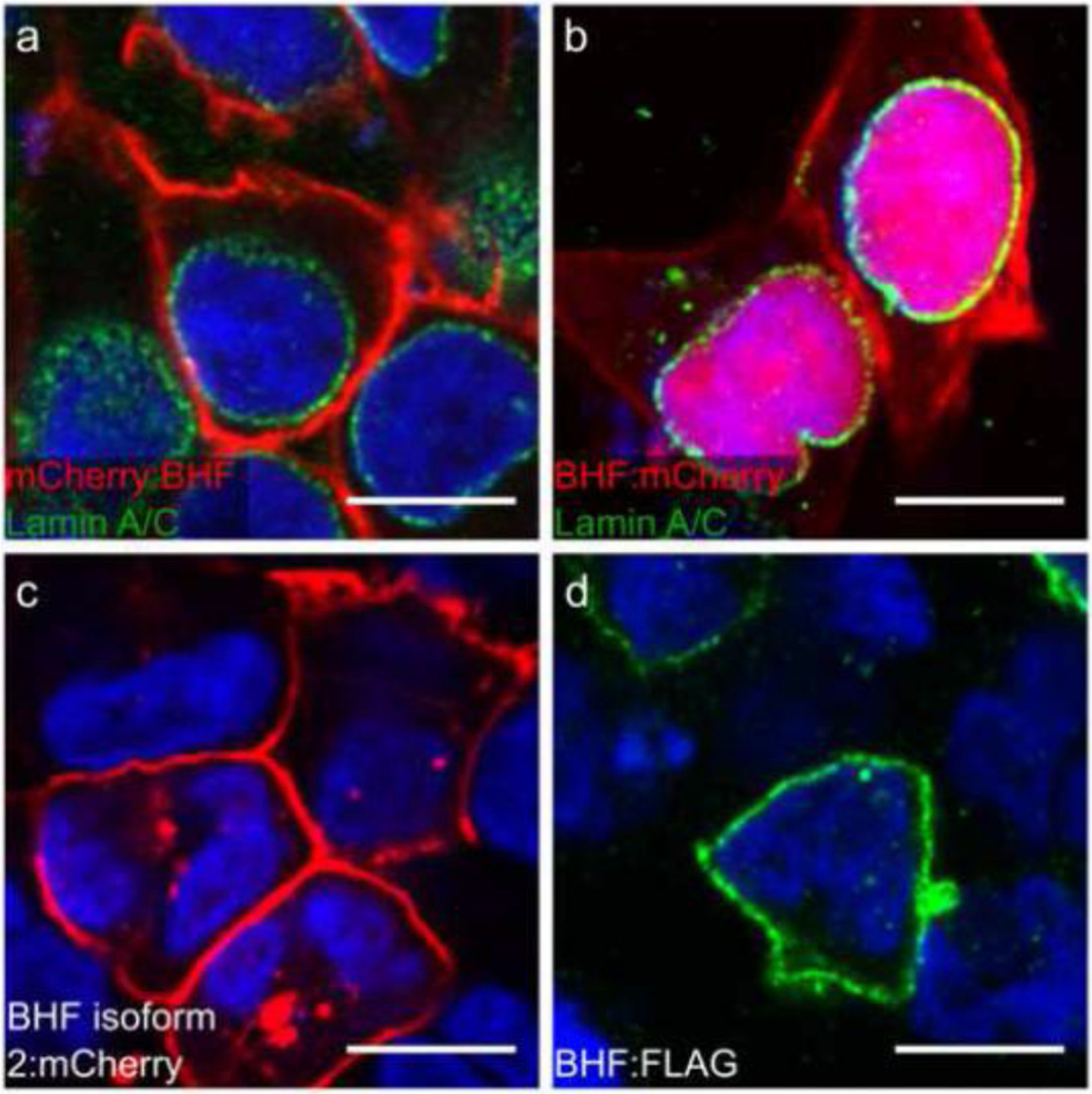


Fig. 7.

Nuclear localization of BHF. Confocal images of (a) mCherry:BHF (red) and stained with a mouse α-Lamin A/C antibody (green); (b) BHF:mCherry (red) with a mouse α-Lamin A/C stain (green); (c) BHF (isoform 2):mCherry; (d) BHF:FLAG detected with a mouse α-FLAG antibody (green). All images were counterstained with DAPI (blue). Scale bar = 10 μm

Table 1

List of cloning primers used in this study. NΔ or CΔ represents which terminal end is being truncated by *x* many amino acids. Primer sequences corresponding to the coding region are capitalized

Construct	**Primer Sequences (5'➔3') (Forward Primer / Reverse Primer)**	**Vector**
bhf (Isoform 1)	gtcgtatcatccgttactgacaagcaag / ccgttaactcagcatgtcacgg	pGEM®-T-Easy
mCherry:bhf (Isoform 1)	ccccccctcgaggcATGGTGCACGATACCGAGCAA / ccccccggatccTTATCTCCTGTTTTGTACAAGGGCCG	pmCherry-C1
mCherry:bhf CΔ52 (Isoform 1)	ccccccctcgaggcATGGTGCACGATACCGAGCAA / ccccccggatccTTACTTCAAATACTCCCCAGCGCTGT	
mCherry:bhf CΔ102 (Isoform 1)	ccccccctcgaggcATGGTGCACGATACCGAGCAA / ccccccggatccTTAGACCGGTTCGTTTTCCATGTCCC	
mCherry:bhf CΔ152 (Isoform 1)	ccccccctcgaggcATGGTGCACGATACCGAGCAA / ccccccggatccTTACTGAGCGGGCTTGCTAGTTTCTG	
mCherry:bhf CΔ202 (Isoform 1)	ccccccctcgaggcATGGTGCACGATACCGAGCAA / ccccccggatccTTACTTCTTCAAATGTAACGCAGTCAGGATTCG	
bhf:mCherry (Isoform 1)	cccccgctagcATGGTGCACGATACCGAGCAATTG / ccccccctcgagTCTCCTGTTTTGTACAAGGGCCGC	pmCherry-N1
NΔ50 bhf:mCherry (Isoform 1)	cccccgctagcATGAGGAGAACTAAACACGATCATCAAAAGCT / ccccccctcgagTCTCCTGTTTTGTACAAGGGCCGC	
NΔ100 bhf:mCherry (Isoform 1)	cccccgctagcATGACCGCGATTTCAAGATTCCCATCA / ccccccctcgagTCTCCTGTTTTGTACAAGGGCCGC	
NΔ150 bhf:mCherry (Isoform 1)	cccccgctagcATGGACCAACTACCCCACGAAATCCT / ccccccctcgagTCTCCTGTTTTGTACAAGGGCCGC	
NΔ200 bhf:mCherry (Isoform 1)	cccccgctagcATGCATCAATGGAAGGGGCAGGGG / ccccccctcgagTCTCCTGTTTTGTACAAGGGCCGC	
bhf:FLAG (Isoform 1)	cccccgcggccgccATGGTGCACGATACCGAGCAATTGC / ctcgagttacttatcgtcgtcatccttgtaaTCTCTCCTGTTTTGTACAAGGGCCG	pCMV-Tag3B
bhf (Isoform 2)	gttgttatcatccgttactgacaagcaag / gatcgattcgtggttggaacttcattt	pGEM®-T-Easy
mCherry:bhf (Isoform 2)	ccccccctcgaggcATGGTGCACGATACCGAGCAA / ccccccggatccTCACACACCATTCGACCTCACCT	pmCherry-C1
bhf:mCherry (Isoform 2)	cccccgctagcATGGTGCACGATACCGAGCAATTG / ccccccctcgagCACACCATTCGACCTCACCTTGC	pmCherry-N1

Table 2

Summary statistics for genes in the *fuhc* locus

Gene	# of Sequences	# of Haplotypes	Haplotype diversity	π (All Sites)	π (Synonymous Sites)	Π (Nonsynonymous Sites)	Θ_w (per site)
bhf	29	13	0.93	0.02	0.07	0.01	0.02
cfuhc^{sec}	16	15	0.99	0.04	0.02	0.04	0.04
cfuhc^{tm}	13	11	0.97	0.03	0.07	0.01	0.03
hsp40-1	14	13	0.99	0.03	0.08	0.01	0.03

Table 3

Results of Tajima's D and Fu and Li's D* and F* neutrality test for *bhf*

	Value: All sites	p value	Value: Nonsynonymous sites	p value	Value: Synonymous sites	p value
D	0.69	0.90	−0.05	0.48	0.97	0.96
D*	0.89	0.94	−0.80	0.16	1.59	0.99
F*	0.98	0.95	−0.66	0.21	1.63	1.00