Developmental and Comparative Immunology 69 (2017) 60–74

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

Balancing selection on allorecognition genes in the colonial ascidian *Botryllus schlosseri*

Marie L. Nydam [a, *], Emily E. Stephenson [a, b], Claire E. Waldman [a], Anthony W. De Tomaso [c]

[a] Division of Science and Mathematics, Centre College, 600 W. Walnut Street, Danville, KY 40422, United States
[b] Centre for Infectious Disease Research, P.O. Box 34681, Lusaka, 10101, Zambia
[c] Department of Molecular, Cellular, and Developmental Biology, University of California Santa Barbara, Santa Barbara, CA 93106, United States

ARTICLE INFO

Article history:
Received 24 November 2016
Received in revised form
22 December 2016
Accepted 22 December 2016
Available online 24 December 2016

Keywords:
Botryllus schlosseri
Allorecognition
Balancing selection
fuhc^{sec}
fuhc^{tm}
Ascidian

ABSTRACT

Allorecognition is the capability of an organism to recognize its own or related tissues. The colonial ascidian *Botryllus schlosseri*, which comprises five genetically distinct and divergent species (Clades A-E), contains two adjacent genes that control allorecognition: *fuhc*^{sec} and *fuhc*^{tm}. These genes have been characterized extensively in Clade A and are highly polymorphic. Using alleles from 10 populations across the range of Clade A, we investigated the type and strength of selection maintaining this variation. Both *fuhc* genes exhibit higher within-population variation and lower population differentiation measures (F_{ST}) than neutral loci. The *fuhc* genes contain a substantial number of codons with >95% posterior probability of $d_N/d_S > 1$. *fuhc*^{sec} and *fuhc*^{tm} also have polymorphisms shared between Clade A and Clade E that were present prior to speciation (trans-species polymorphisms). These results provide robust evidence that the *fuhc* genes are evolving under balancing selection.

1. Introduction

Allorecognition is the ability of an organism to distinguish self from non-self, or related individuals from non-related individuals (kin discrimination). Well-studied systems of allorecognition are present on all branches of the tree of life: from prokaryotes such as Bacteria (Cao et al., 2015; Rendueles et al., 2015; Stefanic et al., 2015; Wenren et al., 2013), to unicellular eukaryotes like Archamoebae (Espinosa and Paz-y-Miño-C, 2012), Ciliophora (Cervantes et al., 2013; Chaine et al., 2010), Dikarya (Brückner and Mösch, 2012; Goossens et al., 2015; Glass et al., 2000; Raudaskoski, 2015) and Mycetozoa (Kalla et al., 2011; Kaushik et al., 2006; Mehdiabadi et al., 2006; and Romeralo et al., 2012), to multicellular eukaryotes such as Plantae (Allen and Hiscock, 2008; Iwano and Takayama, 2012; Pandey, 1960; Wu et al., 2013) and Animalia (Flajnik and Kasahara, 2010, Hughes et al., 2004; Karadge et al., 2015; Puill-Stephan et al., 2012, Raftos, 1994; Saito, 2013; Smith et al., 2012; Taketa and De Tomaso, 2015; Westerman et al., 2009).

The genetic basis of allorecognition is not well understood in invertebrates, with the exception of the hydroid *Hydractinia symbiolongicarpus* and the ascidians *Ciona robusta* and *Botryllus schlosseri* (Grosberg and Plachetzki, 2010). In *C. robusta*, allorecognition loci provide a mechanism for inbreeding avoidance (Ban et al., 2005; Harada et al., 2008), similar to the SI (Self-Incompatibility) loci in angiosperms (Iwano and Takayama, 2012). In *B. schlosseri*, allorecognition loci allow neighboring colonies to determine whether or not they are related; only related colonies fuse to form a new colony.

B. schlosseri comprises five cryptic species: Clades A-E (Bock et al., 2012; Lopez-Legentil et al., 2006); all previous allorecognition work has focused on Clade A only. Clade A is found in all the major oceans of the world (Bock et al., 2012; Yund et al., 2015). Clades B-D have only been described from <10 individuals each, found in 1–2 locations in the western Mediterranean Sea, the southern Bay of Biscay or the English Channel (Bock et al., 2012; Lopez-Legentil et al., 2006). Clade E is widespread in the northeastern Atlantic Ocean and Mediterranean Sea (Bock et al., 2012; Lopez-Legentil et al., 2006). The species within the *B. schlosseri* complex are genetically divergent: 10.8–16.5% at the mitochondrial cytochrome oxidase I locus (mtCOI; Bock et al., 2012). Bayesian clustering of seven microsatellite loci revealed no evidence of gene

* Corresponding author.
E-mail addresses: marie.nydam@centre.edu (M.L. Nydam), emily.stephenson.rotary@gmail.com (E.E. Stephenson), claire.waldman@centre.edu (C.E. Waldman), anthony.detomaso@lifesci.ucsb.edu (A.W. De Tomaso).

http://dx.doi.org/10.1016/j.dci.2016.12.006
0145-305X/

flow between Clades A, D or E (Clades B and C were not analyzed) (Bock et al., 2012).

The Clade A allorecognition determinant was initially thought to be a single gene with two alternative transcripts (De Tomaso et al., 2005) but has since been characterized as two genes separated by 227 base pairs (Nydam et al., 2013a). *fuhc*^{sec} encodes a 555 amino acid protein that is secreted from the cell, and *fuhc*^{tm} encodes a 531 amino acid membrane-bound protein (Nydam et al., 2013a). Both proteins are highly polymorphic, although *fuhc*^{sec} is more variable than *fuhc*^{tm} (Nydam et al., 2013b). Polymorphisms in both proteins predict allorecognition outcome. However, *fuhc*^{sec} may be determining or driving histocompatibility (Nydam et al., 2013a). Evidence for this possibility comes from repeated pairings between two genotypes that result in rejections so weak that they are often mistaken as non-reactions (Nydam et al., 2013a). Rejection severity is thought to be linked to divergence between alleles: more amino acid differences between alleles resulting in stronger rejections (Scofield and Nagashima, 1983). In this weak rejection, the *fuhc*^{sec} alleles were nearly identical but the *fuhc*^{tm} alleles were divergent.

The extensive polymorphism in allorecognition loci is maintained by natural selection, either balancing or directional (reviewed in Nydam and De Tomaso, 2011). Balancing selection maintains polymorphism in a population by pushing an allele to an intermediate frequency, whereas directional selection reduces polymorphism in a population by moving an allele to fixation (Hedrick, 2007). Several population genetic tests are commonly used to detect selection; in some cases these tests can also differentiate between balancing and directional selection. A summary of the tests used in the allorecognition literature and in this study is presented in Table 1.

In several allorecognition systems, balancing selection has maintained polymorphisms that were present prior to speciation (Figueroa et al., 1988; Guo et al., 2011; Ioerger et al., 1990; Sato et al., 2002; Wu et al., 1998). Selection has been shown to operate on Clade A *fuhc*^{sec} and *fuhc*^{tm} when sequences from the eastern Pacific and northwestern Atlantic Oceans were analyzed. ω statistics (d_N/d_S) revealed codons throughout *fuhc*^{sec} and *fuhc*^{tm} with >95% probability of $\omega > 1$ and values of Tajima's D that deviated from neutrality at allorecognition loci but not at any of 12 housekeeping loci (Nydam et al., 2013b). Additionally, a statistically significant lack of population structure at both *fuhc*^{sec} and *fuhc*^{tm} was contrasted with strong structure at 12 housekeeping loci (Nydam et al., 2013b), mtCOI (Lopez-Legentil et al., 2006), and microsatellites (Ben-Shlomo et al., 2010; Bock et al., 2012). However, the type of selection could not be identified with confidence: ω statistics cannot differentiate between balancing and directional selection, Tajima's D statistics provided evidence for directional selection, and analyses of population structure pointed towards balancing selection (Nydam et al., 2013b).

The current study aims to conclusively determine whether balancing or directional selection is acting on Clade A *B. schlosseri* allorecognition loci. In our previous study, we only included Clade A populations from the eastern Pacific and northwestern Atlantic Oceans. This limited sampling could explain the inconsistent patterns of selection inferred in Nydam et al., 2013b. Here, we include Clade A sequences from four new populations (two from the English coast of the English Channel and two from the French coast of the English Channel). We can now analyze selection patterns in 10 populations, which may result in a stronger consensus about which type of selection is most prevalent in the species as a whole. Conversely, populations in different geographic regions may be experiencing different types of selection. Adding a new geographic region allows us to contrast the evolutionary history of Clade A *fuhc*^{sec} and *fuhc*^{tm} in three geographic regions (eastern Pacific, northwestern Atlantic, and English Channel). Inclusion of only Clade A sequences in Nydam et al., 2013b limited the scope of the population genetic analyses we could perform. In the current study, we include Clade E *fuhc*^{sec} and *fuhc*^{tm} sequences to make inferences of selection that were not possible with intraspecific data from Clade A, including tests of trans-species polymorphism.

2. Materials and methods

2.1. Sampling

Clade A and Clade E colonies were collected from floating docks in the eastern Pacific, northwestern Atlantic, Bay of Biscay, Mediterranean, and English Channel from 2009 to 2015 (Table 2). Colonies from Clades B, C and D were not recovered, despite collecting in all locations where they had been previously found. Colonies were collected from several locations in each harbor or marina in order to include a broad representation of the *fuhc* alleles in the population. Single systems were cut from individual colonies and immediately placed in RNAlater® (Thermo Fisher Scientific, Waltham, MA, USA). Samples were stored at −80 °C within 12 days of collection.

2.2. Amplification and sequencing: fuhc genes

Because morphological characters differentiating Clade A and

Table 1
Population genetic tests used in this study to detect directional and balancing selection at *fuhc*^{sec} and *fuhc*^{tm}.

Test	Inference
Neutrality Statistics	Negative values = directional selection Positive values = balancing selection Values = 0 = neutral evolution (no selection)
McDonald Kreitman	Excess of nonsynonymous fixed differences = directional selection Excess of nonsynonymous polymorphisms = balancing selection Ratio of nonsynonymous to synonymous polymorphism is equal to the ratio of nonsynonymous to synonymous divergence = neutral evolution (no selection)
HKA	Difference in amount of polymorphism and divergence between two loci cannot be explained by difference in mutation rate = balancing selection Difference in amount of polymorphism and divergence between two loci can be explained by difference in mutation rate = neutral evolution (no selection)
ω (dN/dS)	Values > 1 = directional or balancing selection Values = 1 = neutral evolution (no selection)
Distribution of polymorphism	Lower polymorphism within populations and higher differentiation between populations than neutral loci = directional selection Higher polymorphism within populations and lower differentiation between populations than neutral loci = balancing selection Similar levels of polymorphism within populations and similar levels of differentiation between populations to neutral loci = neutral evolution (no selection)