The distribution and unexpected genetic diversity of the non-indigenous annelid *Ficopomatus enigmaticus* in California

Alison Yee\(^1\), Joshua Mackie\(^2\) and Bruno Pernet\(^{1,*}\)

\(^1\)Department of Biological Sciences, California State University Long Beach, 1250 Bellflower Blvd, Long Beach, CA 90840, USA  
\(^2\)Gilmac P/L, Suite G04, 1 Havelock Street, West Perth 6005, Australia  

Author e-mails: alisonkyee@gmail.com (AY), joshuamackie@gmail.com (JM), bruno.pernet@csulb.edu (BP)  

*Corresponding author

**Abstract**

The non-indigenous annelid *Ficopomatus enigmaticus* has been established in San Francisco Bay since at least 1921, but in the past 30 years it has also been found in other parts of California. In the summer of 2017 we surveyed 136 sites to determine its current distribution in the state. We found *F. enigmaticus* at 23 sites ranging from San Francisco Bay in the north to Newport Bay in the south. Populations were concentrated in four regions: San Francisco Bay, Monterey Bay, Santa Barbara, and sites in Los Angeles and Orange Counties. Presence sites did not differ systematically in salinity or temperature from absence sites, but all presence sites appeared to have restricted exchange of water with nearby oceanic habitats. Data on the timing of first discovery in each region is roughly consistent with the hypothesis of southward spread of propagules from the San Francisco Bay population. To further test this hypothesis, we obtained mitochondrial DNA sequences from individuals collected from four sites nearly spanning the current latitudinal range of *F. enigmaticus* in California. Recent work from Australia suggests that there is substantial within-population cryptic genetic diversity in *F. enigmaticus*, with individual aggregations containing individuals whose mitochondrial cytochrome B haplotypes fall into one of two very distinct (~ 19% uncorrected genetic distance) haplotype groups, Clades 1 and 2. We found a similar pattern in California, with Clade 1 and Clade 2 individuals co-occurring at two of the four sites we sampled. Three of ten known *F. enigmaticus* haplotypes occurring in Australia were observed in California populations; four haplotypes observed in California have not previously been reported. Analysis of haplotype distributions suggests that central California populations may be derived from the San Francisco Bay population, while unique haplotypes present in the Long Beach population suggest the possibility of a second independent introduction in that region. Additional genetic data from populations of *F. enigmaticus* around the globe are needed to resolve the invasion history and systematics of these widespread serpulids.

**Key words:** cryptic species, polychaete, range expansion, Serpulidae

**Introduction**

The serpulid annelid *Ficopomatus enigmaticus* (Fauvel, 1923) may have originated in temperate regions of the Indian Ocean (including Australia), though this conclusion is not at all certain (reviewed by Dittmann et al. 2009 and Styan et al. 2017). Since the 1920s it has been recognized as a
Table 1. Temporal patterns of the discovery of established populations of *Ficopomatus enigmaticus* in California, at the coarse geographic scale of counties. The exception is San Francisco Bay, whose shoreline is divided up among several counties.

<table>
<thead>
<tr>
<th>Year of 1st observation</th>
<th>Region</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1921</td>
<td>San Francisco Bay</td>
<td>Carlton 1979</td>
</tr>
<tr>
<td>1994</td>
<td>Monterey County</td>
<td>J. Alicea, in Wasson et al. 2001</td>
</tr>
<tr>
<td>2000</td>
<td>Los Angeles County</td>
<td>Cohen et al. 2002, 2005</td>
</tr>
<tr>
<td>2001</td>
<td>Santa Cruz County</td>
<td>K. Wasson, <em>unpub. obs.</em></td>
</tr>
<tr>
<td>2013</td>
<td>Orange County</td>
<td>Obaza and Williams 2018</td>
</tr>
<tr>
<td>2013</td>
<td>San Diego County</td>
<td>Obaza and Williams 2018</td>
</tr>
</tbody>
</table>

Non-indigenous species in many other locations, and it is now established on warm temperate coasts, especially in estuaries, around the world (Dittmann et al. 2009). Establishment of *F. enigmaticus* in new regions changes the local ecology, in part because this suspension-feeding annelid usually occurs in large aggregations (sometimes referred to as reefs). These aggregations alter the physical structure of benthic habitats (Schwindt et al. 2004; McQuaid and Griffiths 2014), with important direct and indirect effects on native benthic communities (e.g., Heiman and Micheli 2010). In aggregations, *F. enigmaticus* occurs at densities on the order of 100,000 individuals per m² (Dittmann et al. 2009), and at such high densities their active suspension feeding also has important effects on the plankton (e.g., Bruschetti et al. 2008; Pan and Marcoval 2014). *Ficopomatus enigmaticus* can also be an economically important fouler of man-made structures (e.g., boat hulls, cooling water intakes) (Dittmann et al. 2009).

In the northeast Pacific, *F. enigmaticus* (although not yet described) was first reported in California in 1921, in San Francisco Bay (Carlton 1979; Cohen and Carlton 1995). It was not found outside of San Francisco Bay until the early-mid 1990s, when a population was discovered ~ 150 km to the south in Elkhorn Slough, Monterey Bay (Wasson et al. 2001). Over the next two decades, populations were discovered scattered along the coast from Point Conception to San Diego (Table 1). To our knowledge, no other species of *Ficopomatus* has been observed in California, though two other species in the genus are known as non-indigenous species on the Pacific coast of Mexico: *F. miamiensis*, and *F. uschakovi* (Bastida-Zavala et al. 2017). As adults, these three species are easily distinguished by opercular characteristics.

Like many accounts of the introduction(s), distribution, and spread of non-indigenous species, the above summary of the history of *F. enigmaticus* in California is imprecise because most discoveries of new populations were made incidentally while observers were doing other work. Other than the studies of Cohen et al. (2002, 2005), which focused on surveys of diverse non-indigenous species in three major port complexes in southern California, and that of Pernet et al. (2016), which included rapid surveys of a small number of southern California sites for the presence of *F. enigmaticus*,
there have been no planned searches for established populations of this species in the state. Such survey data would be very helpful in identifying “new” populations that may need to be managed, and in providing a baseline set of data with which to evaluate future range expansions or contractions.

However, interpretation of the results of any such survey has been complicated by recent findings of unexpected genetic diversity within *F. enigmaticus*. Specifically, Styan et al. (2017) sequenced a 266 base pair fragment of the mitochondrial cytochrome b gene to make inferences on the phylogeography of *F. enigmaticus* across its range in southern Australia. Analyses of these data unexpectedly identified two very distinct haplotype groups (Clades 1 and 2) that on average differed at 19.2% of the bases (Styan et al. 2017) also identified a third haplotype group in southeast Australia, Clade 3, but members of that clade were morphologically distinct from *F. enigmaticus*, and more similar to *F. uschakovi*). Members of Clades 1 and 2 often co-occurred at a very small spatial scale, within the same aggregations. Styan et al. (2017) also presented results of an analysis of dominant nuclear markers (inter-sequence simple repeats) that suggested that there might be concordant differences in the nuclear genome. Together these data suggest the possibility—which requires further evaluation—that the taxon *F. enigmaticus* in fact includes two cryptic species. Thus instead of simply asking “where does *F. enigmaticus* occur?” in any part of its current range, we can now also ask “what is the distribution of haplotypes?” in that region.

In this paper, we begin to address these two questions in California. We first report on a survey of 136 sites carried out in the summer of 2017 for the presence of *Ficopomatus enigmaticus*, from Tomales Bay in the north to Tijuana River in the south. To address the second question, we report on a survey of mitochondrial haplotypes of specimens from a subset of the California sites where *F. enigmaticus* is present, chosen so as to nearly span its latitudinal range in the state. Data on the distribution of *F. enigmaticus* on the west coast of the United States are critical to understanding the trajectory of range expansion and contraction in this species, and data on their genetic diversity are needed to reconstruct the invasion history and systematics of these serpulids.

**Materials and methods**

*Distributional surveys*

In May–August 2017 we conducted surveys for the presence or absence of *Ficopomatus enigmaticus* at 136 sites along the coast of central and southern California. To identify sampling sites prior to surveys, we examined Google Maps satellite views of the California coast from the Tijuana River in the south to Tomales Bay in the north. *Ficopomatus*
enigmaticus usually occurs in wave-protected habitats with hard substrate available and water of salinity < 35 psu (Dittmann et al. 2009). We selected sites in part based on the likelihood that one or more of these conditions was met, and in part based on accessibility. We also examined all prior reports of the distribution of this species in California (including a few personal communications from biologists along the coast), and sampled at many of those sites as well to confirm the presence of established populations. We did not survey north of Tomales Bay because to our knowledge F. enigmaticus has never been reported as present north of San Francisco Bay and because of limits on time and funding. It would be valuable to survey this region, however, in case F. enigmaticus begins (or has already begun) to spread to the north.

Intertidal sites were visited around low tide, but floating docks were visited whenever convenient. At each site, at least one author (Yee) but more commonly two (Yee and Pernet) conducted a visual survey of the intertidal zone, docks, or other anthropogenic structures for ~ 5 min (the time it took to thoroughly examine ~ 10 m of shoreline or structure). Adults of F. enigmaticus produce distinctive white tubes and are usually easily observed, even from a distance. Where rubble or small boulders were intertidal substrate, we examined both their upper and lower surfaces. If F. enigmaticus was present at a site, we preserved a small sample in 95% ethanol. At each site, we measured water salinity and temperature in the top 25 cm of the water column using a YSI Pro 2030. Salinity measurements were periodically checked with a calibrated refractometer, and temperature measurements with an alcohol thermometer.

Genetic diversity

We had intended to study the genetic diversity of F. enigmaticus in California using the ethanol-preserved samples collected in our distributional surveys. However, we had difficulty extracting high-quality DNA from these samples. Thus, in March and April 2018 we collected new samples from four sites nearly spanning the species’ known latitudinal range in California: Lake Merritt, Alameda County (the northernmost site; this is the site from which F. enigmaticus was first discovered in California in the early 1920s: Carlton 1979); Elkhorn Slough, Monterey County; Arroyo Burro Creek, Santa Barbara County; and Long Beach, Los Angeles County (the southernmost site). (These are sites 24, 48, 73, and 99 in Table S1.) At each site, a single chunk of aggregation large enough to contain ~ 50–100 adult worms was collected from the intertidal zone, with one exception: at the southernmost site, Long Beach, small chunks of aggregations were collected from two locations separated by ~ 200 m shoreline distance. We sampled separate locations within this site to begin to identify the scale of spatial population genetic variation in these serpulids. In all cases, worms were returned to the laboratory alive within 24 h of collection, then immediately dissected from tubes and processed for DNA extraction.
We initially extracted genomic DNA from a small piece of the posterior end of each worm using Qiagen DNEasy extraction kits; however, we found that PCR amplifications using this template DNA frequently failed. We then switched to extracting genomic DNA using Chelex 100 resin (Bio-Rad); these extracts consistently yielded good PCR results. We cut off ~ 0.5–1 mm of the posterior end of each living worm with a sterile razor blade, and immediately immersed it in a suspension of 10% Chelex in a 1.5 ml tube. Tubes were briefly agitated with a vortex mixer, then incubated at 95–100 °C for 15–20 min. They were vortexed again, then frozen at −20°C. Immediately before use, tubes were defrosted and centrifuged to separate the supernatant from resin and remaining tissue. The supernatant was used as template in PCR reactions. The rest of each individual living worm was briefly relaxed in a 1:1 mixture of seawater and 7.5% MgCl₂, then fixed in 5% formalin in seawater.

In order to generate a dataset easily comparable to that of Styan et al. (2017), we used PCR to amplify a fragment of the mitochondrial gene Cytochrome Oxidase B using the primers CytB 424F (5’-GGW TAY GTW YTW CCW TGR GGW CAR AT-3’) (Boore and Brown 2000) and cobr825 (5’-AAR TAY CAY TCY GGY TTR ATR TG-3’) (Burnette et al. 2005). We carried out amplifications in 50 μl reaction volumes containing 1X ThermoPol buffer (New England Biolabs; this buffer includes 2 mM MgSO₄), 0.2 mM of each dNTP, 1 μM of each primer, 1.25 U Taq DNA polymerase (New England Biolabs), and 4 μl of template DNA. The thermal cycling profile was 1 min at 95 °C, 40 cycles of 30 sec at 95 °C, 30 sec at 48.6 °C, and 45 sec at 68 °C, then an additional extension at 72 °C for 5 min. We purified amplified products using Qiagen QiaQuick PCR purification kits, following the manufacturer’s instructions.

Sequencing was conducted in both directions by Eurofins Genomics (Kentucky, USA). Chromatograms were examined and trimmed using 4Peaks (A. Griekspoor and T. Groothuis, nucleobytes.com), and forward and reverse sequences for each individual assembled into contigs and checked by eye using CLC Main Workbench (Qiagen.com). In this fashion we obtained 264 bp of high-quality Cyt B sequence for 76 individuals. We aligned our sequences with 40 of the 48 *F. enigmaticus* Cyt B sequences analyzed by Styan et al. (2017; GenBank KP863736–KP863777). We omitted eight of their sequences from analysis because they were shorter than our 264 bp sequences, or did not overlap with them perfectly.

Analyses of genetic data from these 116 individuals were carried out following the methods of Styan et al. (2017) to facilitate comparison to that study. Phylogenetic analyses were carried out with MrBayes ver. 3.1.2 (Ronquist and Huelsenbeck 2003) with the same substitution models and run parameters used by Styan et al. (2017). Pairwise distances (both uncorrected distances, and Tamura-Nei distances) among the major clades of *F. enigmaticus* were calculated with MEGA 7 (Kumar et al. 2016).
Analysis of Molecular Variance (AMOVA) was carried out with Arlequin 3.5.2 (Excoffier and Lischer 2010), using Tamura-Nei distances. The AMOVA was also run using only frequencies of shared haplotypes (a weighting of 1, between all haplotypes) to characterize shuffling pattern without the effect of mutational divergence. Ten sites—four in California, and six in Australia (excluding clade 3, individuals at Yowaka River and Wallagaraugh River, east of Bass Strait)—were analyzed. The covariance component $F_{CT}$ refers to differences between Australia and California (USA), $F_{ST}$ to between site differences, and $F_{SC}$ to between sites within a country. Significance of $F$-statistics was calculated by 10,000 permutations.

Results

Distribution

We found populations of *Ficopomatus enigmaticus* at 23 of the 136 sites surveyed (Figure 1; Table S1). The northernmost populations were in San Francisco Bay (where this species is widespread); the southernmost populations were in Newport Bay. We found living *F. enigmaticus* at all of these 23 sites except for four in Santa Cruz County (sites 39, 43, 44, and 46; Table S1). During our 2017 surveys we found numerous empty tubes of *F. enigmaticus* at these four sites, but no living worms. This was unusual – at all other sites, we found either living worms or no traces (e.g., empty tubes) of *F. enigmaticus* at all. We returned to these four Santa Cruz County sites in August 2018, and at that time observed living worms at all four of them.

To our knowledge, nine of the 23 populations we found have not been described before in the literature. These include five populations in Santa Cruz County (sites 39, 43, 44, 46, and 47) and four in Santa Barbara County (sites 72, 76, 77, and 79). Though records of these populations were not available in the literature, biologists in these areas have been aware of some of them for some time. For example, Dr. Kerstin Wasson (*pers. comm.*) identified *F. enigmaticus* from site 39 (San Lorenzo River) in 2001.

Our predictions about site characteristics made prior to surveys were imperfect, and sites in fact varied greatly in salinity, temperature, and substrate type. Hard substrate was available at all sites, though it varied in substantially in percent cover (which we did not measure). Sites where *F. enigmaticus* were present did not appear to differ systematically in salinity or temperature from sites where it was not present (Figure 2). Particularly striking was the range of environmental salinities this serpulid appeared to tolerate – we found living *F. enigmaticus* at sites with salinities as low as 1.4 psu, and as high as 37.1 psu (median = 20.69 psu). The populations we observed were always in relatively wave-protected habitats, and always in habitats that appeared to have partially or completely restricted exchange of water with the nearby ocean. Sometimes these sites were quite small in terms of suitable habitat. For example, sites 46, 73, 76,
Distribution and diversity of *Ficopomatus enigmaticus* in California


Figure 1. Sites surveyed for the presence of *Ficopomatus enigmaticus* in central and southern California in the summer of 2017. Sites are indicated by circles. Open circles are sites at which no *F. enigmaticus* were found, and filled (red) circles are those at which *F. enigmaticus* were found. The numbers adjacent to sites at which *F. enigmaticus* were found correspond to site codes in Table S1. Numbers in red (24, 48, 73, and 99) indicate the four populations from which samples were collected for analyses of genetic diversity in 2018.

and 79 were the mouths of small creeks that ended at a beach, closed off from the ocean by wide sand berms, at least at the time we surveyed them.

**Genetic diversity**

We obtained Cyt B sequences from 76 individuals of *F. enigmaticus* from four sites along the California coast: Lake Merritt, Alameda County (AL, 16 individuals), Elkhorn Slough, Monterey County (MO, 12), Arroyo Burro, Santa Barbara County (SB, 17), and Long Beach, Los Angeles County (LA, 31; at this site, 17 individuals were collected from one aggregation, and 14 from another ~ 200 m distant). These are sites 24, 48, 73, and 99, respectively.
Figure 2. Water temperature plotted against salinity for the 136 survey sites, including both sites at which *Ficopomatus enigmaticus* was not found (open circles) and those at which it was found (filled circles).

(Table S1). The 76 sequences (Genbank MK334048–MK334123) included seven haplotypes. We also analyzed Cyt B sequences from 40 individuals studied by Styan et al. (2017); these fell into 10 haplotypes. Three haplotypes (haplotypes C, E, and H in Styan et al. (2017)) were found in both California and Australia. The pooled data included a total of 14 unique haplotypes.

Bayesian analysis of the 116 sequences and three outgroup sequences yielded a phylogram very similar to that described by Styan et al. (2017) (Figure 3). A small group of individuals from far-eastern Australia shared a single haplotype and clustered with *F. macrodon*; Styan et al. (2017) argued that these were likely misidentified as *F. enigmaticus* and in fact were either *F. uschakovi* or members of a closely related undescribed species. Almost all remaining sequences fell neatly into one of two major clades, called Clades 1 and 2 by Styan et al. (2017). In our analyses, two clades nearly identical to those of Styan et al. (2017) were recovered, each including individuals from both California and Australia. Because the clades we recovered were nearly identical to those of Styan et al. (2017), we also call them Clades 1 and 2. The position of one Cyt B haplotype (found in a Californian individual from Lake Merritt [ALLM 11] and an Australian individual from Hopkins River [Ho4]) was not resolved in our phylogenetic analysis, where that haplotype forms a polytomy with Clades 1 and 2. In Styan et al.’s (2017) analysis, that haplotype clustered in Clade 1. Unless otherwise specified, references to Clades 1 and 2 below refer to the haplotype groups recovered in our analyses (Figure 3).
As expected, genetic distances between clades were very similar to those reported in Styan et al. (2017). In our analyses, the mean uncorrected distance between sequences in Clade 1 and Clade 2 was 19.6% (Table 2). All other comparisons between clades showed greater genetic distances.

There was substantial variation in haplotype frequency among the four California sites (Table 3). We found four haplotypes in the northernmost site, Lake Merritt. Two of these clustered in Clade 1 (C and N); one in Clade 2 (H); and one was not assigned to a major clade (E). At each of the two central California sites (Elkhorn Slough and Arroyo Burro Creek), we found only two haplotypes, C and N. These were both in Clade 1 and represented a subset of the haplotypes found in Lake Merritt. The sample from Long Beach included four haplotypes. One of them was C, the Clade 1 haplotype found in the three northern populations, but the other three
Table 3. Distribution of Cyt B haplotypes across sampling sites in California. The two Long Beach locations were separated by ~200 m. The three haplotypes labeled with asterisks were named in Table 1 of Styan et al. (2017) and are shared with individuals from southern Australia; the four remaining haplotypes are unique to worms sampled from California. Haplotypes are also labeled (letters in circles) in Figure 3. Note that Haplotype E was not assigned to a clade in our analysis, but in that of Styan et al. (2017) it was assigned to Clade 1.

<table>
<thead>
<tr>
<th>Clade</th>
<th>Haplotype</th>
<th>1</th>
<th>2</th>
<th>not assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lake Merritt (AL)</td>
<td>C</td>
<td>L</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>Elkhorn Slough (MO)</td>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arroyo Burro (SB)</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long Beach (LA)</td>
<td>14</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location 1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Location 2</td>
<td>5</td>
<td>2</td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

(two in Clade 1 [L, M] and one in Clade 2 [O]) were unique to the Long Beach population. Haplotypes of worms from the two Long Beach locations (separated by ~ 200 m) were similar except that the smaller sample of individuals at Location 2 did not include one rare haplotype found at Location 1 (Table 3).

Haplotype C was found at all four California sites, and was also widespread in southern Australia (Styan et al. 2017). Two of the other Lake Merritt haplotypes, E and H, were also shared with specimens in southern Australia.

In AMOVA of Clade 1 and 2 Cyt B sequences, there were significant (P < 0.001) differences in haplotype Tamura-Nei distances among sites as a whole and among sites within countries ($F_{ST} = 0.392$, $F_{SC} = 0.430$), but there was no difference between countries ($F_{CT} = -0.066$, $P = 0.781$). AMOVA using haplotype frequencies (referred to as conventional $F$-statistics in Arlequin) produced a similar pattern of significant distances among sites generally ($F_{ST} = 0.301$, $F_{SC} = 0.301$, $P < 0.001$), and no difference between countries ($F_{CT} = 0.0008$, $P = 0.532$).

Discussion

The survey data described here provide a snapshot of the distribution of *Ficopomatus enigmaticus* on the coast of central and southern California in 2017. In addition to confirming its presence at many sites at which it was already known, we identified nine previously unknown populations in Santa Cruz and Santa Barbara Counties. We did not observe *F. enigmaticus* at sites we surveyed in San Diego County, but Obaza and Williams (2018) observed it at two sites in San Diego Bay in 2013 and 2014 (A. Obaza, pers. comm.). Overall, *F. enigmaticus* appears to be widespread in three major regions of the state: San Francisco Bay, Monterey Bay, and southern California (Santa Barbara, Los Angeles, Orange, and San Diego Counties). The only region of central and southern California where there is currently no evidence of its presence is the central coast between Monterey Bay and Point Conception (Table S1; Figure 1). Data on the past and current
distribution of *F. enigmaticus*, the characteristics of occupied sites, and the genetic diversity of populations in California and Australia allow us to make some initial inferences on how the colonization of California by this species has occurred.

**Characteristics of invaded sites**

Substantial work has been done around the world to link occurrence of *F. enigmaticus* to environmental parameters, especially temperature and salinity (reviewed by Dittmann et al. 2009). These studies suggest that the species can establish reproductive populations at temperatures \( \geq 10 \, ^\circ\text{C} \) and over a large range of salinities (from at least 10–30 psu). In California, we found populations of *F. enigmaticus* present over a wide range of temperatures, and at salinities ranging from 1.4–37.1 psu (Figure 2). Our data are limited in at least two ways, however. First, they represent snapshots of environmental conditions. More helpful would be longer-term records of temperature and salinity data from a given site, to know how mean and extreme values affect the survival of *F. enigmaticus* and its ability to reproduce. Second, we did not assess reproductive ability of worms from each location, so cannot tie that to even instantaneous measures of physical conditions. Such data would be helpful in making better predictions about the types of habitats in which *F. enigmaticus* can settle, grow, and reproduce.

We can make one tentative inference about the effect of salinity on the persistence of *F. enigmaticus* populations based on our observations of the four northern Santa Cruz County populations we observed (sites 39, 43, 44, and 46). In the summer of 2017, despite extensive searching, we found only empty tubes of *F. enigmaticus* in shallow water at these sites. In the summer of 2018, however, shallow habitats at all four sites were occupied by dense aggregations of living worms. We suspect that unusually high rainfall levels in Jan and Feb 2017 (California Nevada River Forecast Center, [https://www.cnrfc.noaa.gov/](https://www.cnrfc.noaa.gov/), accessed 8 Oct 2018) lowered salinities in surface waters at these sites for substantial periods of time, killing *F. enigmaticus* aggregations in shallow water. The presence of living *F. enigmaticus* at these sites in the summer of 2018 might be attributed to recolonization from other sites by larval transport, or, more likely, by recolonization of shallow water habitat by larvae produced by worms that survived the winter freshening in deeper portions of these habitats. The driving of local mortality by reduction in salinity has been hypothesized before for other regions (e.g., Dittmann et al. 2009). It seems unlikely that similar environmental forcing led to a local reduction in density or even extinction of San Diego Bay populations of *F. enigmaticus* present in 2013 and 2014 (Obaza and Williams 2018), but apparently absent in 2017 (this study). Instead it seems more likely that we simply did not examine exactly the same sites that Obaza and Williams (2018) examined. Site coordinates are not listed in their paper.
Hard substrate was available at all sites we sampled, though at some sites it was rare. However, prior work has shown that very large aggregations of *F. enigmaticus* can form at sites where hard substrate is rare and limited to pebbles, the shells of mollusks, or litter such as glass bottles. In such cases, larvae settle on these “nuclei” and build their calcareous tubes, which then serve as settlement substrate for additional juveniles (Schwindt et al. 1999; Schwindt and Iribarne 2000). Thus the availability of hard substrate seems unlikely to be an important factor limiting the distribution of *F. enigmaticus*.

Our data reveal an interesting relationship between another environmental parameter and the occurrence of *F. enigmaticus*: all of the populations we observed were in locations where water exchange with adjacent oceanic habitat was partially or completely restricted (Table S1). Particularly striking was the frequent presence of *F. enigmaticus* in coastal lagoons and “closed” stream mouth estuaries (those separated from the ocean by sand berms that prevent tidal exchange: Jacobs et al. 2010). This is not a novel generalization – Dittmann et al. (2009) noted that *F. enigmaticus* “…thrives in enclosed areas such as estuaries, harbours (wet docks), tidal marshes, channels and lagoons…” In such populations, planktonic larvae may be more likely to be retained near their parents. Gregarious settlement of these larvae might then lead to rapid establishment of easily observed aggregations.

This scenario leads to an obvious question: how are these restricted habitats initially colonized? Some of them (e.g., Elkhorn Slough, San Diego Bay) are near areas where there is substantial traffic by recreational or commercial shipping vessels; in these cases it is possible that populations became established after transport on boat hulls. Other habitats, however, are not navigable, are not particularly close to major boat transit centers, and are truly isolated from the ocean most of the time (e.g., site 73, Arroyo Burro Creek in Santa Barbara County). In such cases colonization seems more likely to have occurred by planktonic larvae during brief periods when the estuary was open to the ocean (e.g., immediately after winter high flow periods; Jacobs et al. 2010). However, we know little about the frequency of occurrence of either adults on boat hulls (but see Davidson et al. 2010 and Zabin et al. 2014) or larvae in the plankton. New tools have recently been developed to identify the presence of *F. enigmaticus* via environmental DNA (Munoz-Colmenero et al. 2018). In future such tools could be used to track the temporal and spatial distribution of larvae around both source and potentially invasible habitats.

*Temporal changes in the distribution of Ficopomatus enigmaticus in California*

Inspection of the time of first discovery of *F. enigmaticus* in different regions of California shows an interesting pattern (Table 1), albeit one that is difficult to interpret clearly given the unknown (and sometimes large)
time lag between establishment and detection of populations (Crooks 2005). Soon after its discovery in Lake Merritt, Oakland, in 1921, *F. enigmaticus* was observed in other parts of San Francisco Bay (Carlton 1979; Cohen and Carlton 1995). It was not observed elsewhere in the state until 1994, when it was discovered in Elkhorn Slough, in the central part of Monterey Bay (Wasson et al. 2001). In the next 20 years it was found in the San Lorenzo River in northern Monterey Bay, and throughout southern California. The last two locations it was identified in the state were in the far south, in Orange and San Diego Counties.

One interpretation of these changes is that after its introduction and establishment in San Francisco Bay, there was “intraregional” transport of *F. enigmaticus* from San Francisco Bay, perhaps in a stepping-stone fashion. For example, it is possible that sometime in the early 1990s, sites in Monterey Bay were colonized, then soon after that sites in Santa Barbara, Los Angeles and then Orange and San Diego Counties. Because dates of first discovery are not necessarily accurate indicators of the time of first establishment (Crooks 2005), this scenario could be consistent with the record of observations in Table 1. One unexplained aspect of this scenario is why it took ~70 years for *F. enigmaticus* to colonize sites outside of San Francisco Bay (but note that lags in range expansion are not uncommon: Crooks 2005). A prediction of this scenario is that California populations to the south of San Francisco Bay are similar to the San Francisco Bay population in terms of genetic diversity, or that they possess only a subset of the diversity present in the San Francisco Bay population.

An alternative possibility is that instead of other California populations being derived from the established San Francisco Bay population by intraregional transport, there have been multiple invasions along the coast from international source populations. *Ficopomatus enigmaticus* is widespread globally (Dittmann et al. 2009), and there is a great deal of international ship traffic not only to San Francisco Bay, but also to the ports of Los Angeles, Long Beach, and San Diego. It is possible that *F. enigmaticus* in southern California are derived from a second international invasion independent of the 1920s invasion of San Francisco Bay. A prediction of this scenario is that California populations to the south of San Francisco Bay might differ from those in San Francisco Bay in terms of genetic diversity, perhaps even including completely different alleles.

The two scenarios mentioned above are not exclusive. The current distribution of *F. enigmaticus* in central and southern California might be the result of both multiple introductions and intraregional transport.

**Inferences on invasion history from genetic data**

As suggested above, genetic data might help to distinguish among alternative invasion scenarios (Geller et al. 2010). However, it is difficult to trace the origins or spread of a non-indigenous species with molecular
markers if there has been limited sampling. This is certainly true of \textit{F. enigmaticus}; we sampled only 12–31 (mean = 19) individuals from each of four sites. Genetic data from an even smaller sample (4–8 individuals from each of nine sites) are all that is available from Australia (Styan et al. 2017). No comparable population genetic data are available from \textit{F. enigmaticus} in any other region of the world. Additional sampling from California, Australia, and other populations of this species around the world would be extremely helpful in making stronger tests of hypotheses. In addition, the primers used in both regions yielded sequence from only a short (264 bp) fragment of the mitochondrial Cyt B gene; data from additional genetic markers would of great utility, as well.

Despite these limitations, the Cyt B data we provide here and that provided by Styan et al. (2017) allow us to make some inferences about both the original introduction of this species into California and subsequent colonization of new sites in the state. First, we were surprised to discover that members of both haplotype clades 1 and 2 are present in both Australia and California (Figure 3). Indeed, AMOVA analyses showed no significant differences between those two geographic regions. Further, in both regions, individuals from the two clades often co-occur in the same small aggregations of tubes. Such variation is consistent with a number of invasion scenarios. One is that individuals of both clades were introduced to California from Australia by 1921, presumably via ballast or hull fouling. Another is that there have been multiple (perhaps many) introductions to California from Australia, and that eventually members of both clades were transported to the state. It is of course also possible that \textit{F. enigmaticus} is not native to southern Australia, and that the genetic diversity there and in California reflect introduction of worms (in one or multiple events) from an unidentified source region. More extensive sampling of \textit{F. enigmaticus} populations globally is clearly needed if we are to have any chance of distinguishing among these hypotheses.

Our data are consistent with the idea that most populations of \textit{F. enigmaticus} in central California were established by propagules from the original San Francisco Bay population. Specifically, the Elkhorn Slough and Arroyo Burro samples (of 12 and 17 individuals, respectively) appear to include only a subset of the mitochondrial haplotypes found in the Lake Merritt population. This is consistent with the hypothesis that a small set of individuals originally from San Francisco Bay invaded these new sites. Larvae of \textit{F. enigmaticus} have a relatively short minimum planktonic duration when reared at natural food concentrations (~ 7–9 d to competence at 16–22 °C; A. Yee, unpubl. data), so rather than occurring via larval dispersal, it seems more likely that \textit{F. enigmaticus} from San Francisco Bay were moved to Monterey and Santa Barbara on the hulls or in ballast of ships. There is substantial traffic of both fishing and recreational vessels between these sites (Wasson et al. 2001; Zabin et al. 2014).
The distribution of haplotypes in the Long Beach population was different from those of the three more northern populations. The Long Beach population shared one haplotype with the Lake Merritt population (and the intermediate two populations), though that haplotype (C) occurred at high frequency in the three northern populations and at low frequency in Long Beach. The other three haplotypes present in Long Beach were private to that population. One possible explanation of this distribution is that the Long Beach population was founded independently of other California populations by a second introduction from an international source (perhaps Australia, where the C haplotype is also geographically widespread and common: Styan et al. 2017). A second introduction in the Long Beach area is plausible, as it is adjacent to two major ports (the ports of Long Beach and Los Angeles) that receive an immense amount of commercial shipping traffic each year. An alternative is that the Long Beach population includes haplotypes from two sources: the C haplotype from San Francisco Bay, and three haplotypes from a second international source. Distinguishing between these hypotheses will require additional sampling in California, as well as globally.

Systematics of Ficopomatus

Additional genetic data are needed to resolve another fundamental question in the biology of Ficopomatus: do the two major mitochondrial clades (1 and 2) detected on both the coasts of California and Australia represent a single biological species with unusually high intraspecific variation at this mitochondrial locus, or do they represent two biological species that have been co-introduced, or sequentially introduced, to California? Studies using nuclear genetic markers are needed to help distinguish between these two hypotheses. To date, the only published analyses of nuclear genetic data are the inter-sequence simple repeats data of Styan et al. (2017); these suggest that members of the two clades differ not only in their mitochondrial genomes, but also in their nuclear genomes, suggesting limited gene flow between members of the two clades. Additional analyses that could yield insight into this question include close examination of the morphology of molecularly-identified specimens; studies of reproductive timing in F. enigmaticus from Clades 1 and 2; or experimental studies of fertilization and development in crosses within and between the two clades.

Acknowledgements

We thank Bengt Allen and Christine Whitcraft for advice on survey design, Philip Spektor and Helen Nguyen for help in surveying sites in Los Angeles County, Madison Thomas and Echelle Burns for help in mapping, Cooper Bahr and Jessica Peria for help with molecular analyses, and Erika Holland for allowing us to carry out molecular work in her laboratory. Kerstin Wasson kindly provided unpublished data on the presence of F. enigmaticus in the San Lorenzo River. We thank Christine Whitcraft and three anonymous reviewers for constructive comments on an earlier draft of the manuscript. This work was supported by grants from the Los Angeles Rod and Reel Club, Southern California Tuna Club, and Reish Student Research Grant in Marine Biology to A.Y., as well as awards from CSU COAST (RR-2016-02) and the National Science Foundation (DEB-1257355) to B.P.
References


Distribution and diversity of *Ficopomatus enigmaticus* in California


**Supplementary material**

The following supplementary material is available for this article:

**Table S1.** Sites surveyed for the presence of *Ficopomatus enigmaticus*, and several of their physical characteristics.

This material is available as part of online article from:

http://www.aquaticinvasions.net/2019/Supplements/A1_2019_Yee_et al_Table_S1.xlsx