UNIVERSITY OF CALIFORNIA
SANTA CRUZ

POPULATION GENOMICS OF THE THREE-SPOT DAMSELFISH (*D. TRIMACULATUS*) SPECIES COMPLEX

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECOLOGY AND EVOLUTIONARY BIOLOGY

by

May B. Roberts

December 2022

The Dissertation of May B. Roberts is approved:

_____
Professor Giacomo Bernardi, Chair

_____
Professor John Thompson

_____
Professor Beth Shapiro

_____
Professor Brian Simison

_____
Professor Michelle Gaither

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

_____
Professor Moisés Bernal

Table of Contents

List of Tables and Figures

Chapter 1:

Chapter 2:

Chapter 3:

Dissertation Abstract

Population Genomics of the three-spot damselfish
(*D. trimaculatus*) species complex

May B. Roberts

Damselfishes (Family: Pomacentridae) are a group of ecologically important, primarily coral reef fishes that include over 400 species. Damselfishes have been used as model organisms to study recruitment (anemonefishes), the effects of ocean acidification (spiny damselfish), population structure and speciation (*Dascyllus*). The genus *Dascyllus* includes a complex of relatively larger bodied species, the *Dascyllus trimaculatus* species complex. This complex is comprised of several species including *D. trimaculatus* itself. The three-spot damselfish, *D. trimaculatus* is a widespread, common, and well-studied coral reef fish species found across the tropical Indo-Pacific that has served as a model species for coral reef fish research. In this dissertation, we expand on what is known of population dynamics of the complex while providing valuable genomic resources to support future research into this model system.

In Chapter 1 of this dissertation, we explore populations at the margins of a species' distribution which tend to be smaller, more isolated, and importantly, exist in habitats at the edge of the species' physiological limit. These conditions make such populations particularly vulnerable to extirpation, especially amid accelerating environmental changes. We focus on population structure based on mitochondrial DNA markers and inferred dynamics of an understudied, potentially vulnerable,

population of an otherwise well studied and widespread coral reef fish. We present phylogeographic results using the most robust set of samples collected in the Ryukyu Archipelago and gain insight into edge population dynamics by comparing to another more well-studied edge population in French Polynesia. We found that despite its proximity to the Coral Triangle and position along the Kuroshio Current, the Ryukyu population of *D. trimaculatus* seems to be a relatively young and closed, suggesting that it may be more vulnerable to extinction than might be otherwise expected.

In Chapter 2 we present the first genome assembly of *Dascyllus trimaculatus*. This assembly contains 910 Mb, 90% of the bases are in 24 chromosome-scale scaffolds, and the BUSCO score of the assembly is 97.9%. Our findings confirm previous reports of a karyotype of 2n = 47 in *D. trimaculatus* in which one parent contributes 24 chromosomes and the other 23. We find evidence that this karyotype is the result of a heterozygous Robertsonian fusion. This assembly was a necessary resource for research in Chapter 3 and will be a valuable resource in the population genomics and conservation of Damselfishes, and continued studies of the karyotypic diversity in this clade.

In Chapter 3, we use low coverage whole genome sequencing to examine the population structure of its Pacific distribution, and again, focus on dynamics of the Ryukyu Archipelago in southern Japan. We find evidence of clear divergence between the Pacific and Indian Ocean populations as well as between the Pacific and the restricted edge population in the Society Islands at the eastern edge of its range. We see a clear signal of introgression of one species within the complex, *D.*

auripinnis described from the Line and Phoenix islands into *D. trimaculatus* that extends across to the west Pacific - further than previously reporte*d.* We also find that there is directional but restricted gene flow of Philippine genotypes into the Ryukyu Archipelago. While we have not yet identified specific isolation and speciation mechanisms, we propose some likely scenarios.

This dissertation builds upon previous research of this complex which had made it a model system in its own right and adds genomic resources to ensure that we continue to further our understanding of this important model system in this genomic age. There are now resources for endless directions and questions to follow regarding this species complex and evolutionary processes.

Understanding mechanisms of dispersal, the role of selection, and the capabilities of local adaptation in a species with relatively restricted gene flow are some of the exciting questions that this project has uncovered, thus placing *Dascyllus trimaculatus* as an ideal model species to address fundamental questions in the ecology and evolution of coral reef fishes.

Acknowledgements

There really is no way I can fully articulate the depth of my gratitude to all the people who made this possible. Thank you, to my advisor Giacomo Bernardi for keeping it light and, in particular, for bringing me in to the OPOR group. I thank the committee members, for their support and advice as my research took its twists and turns. I am blessed with a network of several groups of people, without whom I would not have completed this program. It was the unconditional encouragement, emotional support, financial support, and communication of the belief in me and of my worth by my community that kept me going. My MSc advisor, Mike Berumen for supporting my writing retreat at KAUST to wrap up this dissertation and giving me the much-needed space and freedom to do so at a pivotal time. To my family (Okaasan, Dad, Ken, Ashley, Omama, and fam in the U.S. and Japan), friends (in particular, my core SC grad friends Remy, Nora, Mel, Britt, Ellen – you all have been there every step of the way), and partner (Nanu – you are my rock and safe harbor), thank you, you all made this PhD possible for me. Thank you to the OPOR crew and friends I've met as a guest to the Outer Islands communities. What a life changing collaboration. I will never forget the lessons I've learned and am continuing to learn from our conversations and work together. It's been truly inspiring and eye-opening. To all the undergrads and Ulithi Youth program participants I've worked with, what an honor to be a part of your lives – the kindness, passion, and care for others and the environment by your generation gives me hope and inspires me to keep working for a better world.

Chapter1

The Ryukyu Archipelago as a recent peripheral population of the three-spot damselfish, *Dascyllus trimaculatus*

May Roberts[1,2,*], Iori Kawamura[2], Shiori Kunihiro[2], Masaru Mizuyama[2], Boo Wee Hin[2,3], James Reimer[2,4], Giacomo Bernardi[1]

1. Department of Ecology and Evolutionary Biology, University of California Santa Cruz, 115 McAllister Rd, Santa Cruz, California, USA

2. Graduate School of Engineering and Science, University of the Ryukyus, Senbaru 1, Nishihara, Okinawa, 903-0213, Japan

3. Universiti Kebangsaan, 43600 Bangi, Selangor, Malaysia

4. Tropical Biosphere Research Center, University of the Ryukyus, Senbaru 1, Nishihara, Okinawa, 903-0213, Japan

Abstract

Populations at the margins of a species' distribution tend to be smaller, more isolated, and importantly, exist in habitats at the edge of the species' physiological limit. These conditions make such populations particularly vulnerable to extirpation, especially amid accelerating environmental changes. In this study we focus on structure and inferred dynamics of an understudied, potentially vulnerable, populations of an otherwise well studied and widespread coral reef fish the three-spot damselfish (*Dascyllus trimaculatus)*. We present phylogeographic results using the most robust set of samples collected in the Ryukyu Archipelago and gain insight into edge population dynamics by comparing to another more well-studied edge population in French Polynesia. We found that despite its proximity to the Coral Triangle and position along the Kuroshio Current, the Ryukyu population of *D. trimaculatus* seems to be a relatively young and closed, suggesting that it may be more vulnerable to extinction than might be otherwise expected.

Introduction

Understanding the mechanisms of adaptive evolution and their roles in species' abilities to persist though changing selective pressures is an ever-increasing necessity in the face of rapidly changing environments. As wildlife and their habitats are progressively more affected by human influence through climate change, development, and exploitation, there is growing literature on the consequences of these pressures (Palumbi 2010; Graham et al. 2015; Dulvy, Sadovy, and Reynolds 2003; Doney et al.

2012; Poloczanska et al. 2013; Dulvy, Sadovy, and Reynolds 2003). Species must either shift their ranges to track favored conditions, adapt to new conditions, or often face extinction (Urban 2015; Dulvy, Sadovy, and Reynolds 2003; Parmesan and Yohe 2003).

Populations that are at the margins of a species' distribution, peripheral populations, are often particularly vulnerable to extirpation and at the forefront of environmental changes. This makes them a natural system to study response to changing conditions. Such populations tend to be smaller, more isolated, and importantly, persist in habitats (at least temporarily) that are different than the "ideal" habitats and in fact often at the edge of the species' physiological limit. Compared to species distribution centers, edge populations experience a different swath of natural selection pressures, a more pronounced effect of genetic drift and low gene flow or connection - processes which drive genetic differentiation. These edge populations have been flagged as potentially important sources of genetic diversity including the adaptive variation necessary for species resilience to changing conditions and boundaries (Allendorf et al., 2010; Roberts et al., 2002). They are also important locations for furthering our understanding of evolutionary processes because their persistence is often the result of local adaptation and natural selection. The peripheral system is also interesting when there is high potential for gene flow to counter-act the processes driving genetic differentiation and local adaptation. Marine systems are unique in that most species have high dispersal potential in an environment largely

unrestricted by physical boundaries, but with high variation in habitat availability and structure.

Over the last decades, the three-spot damselfish (*Dascyllus trimaculatus*) species complex has been the subject of several studies aiming to understand its population structure across the Indo-Pacific, the mechanisms driving divergence, and the process of speciation. The complex includes six lineages with restricted geographic ranges, some without taxonomic recognition (see Fig 1). *Dacsyllus trimaculatus* sensu stricto itself has the broadest range, extending from the Red Sea, where it was first described (Ruppell, 182), across the tropical and subtropical Indo-Pacific. There is substantial evidence from phenotypic differences to mitochondrial and genomic data showing a deep divergence between *D. trimaculatus* in the Indian and Pacific Oceans with an active hybrid zone in the eastern Indian Ocean at Christmas Island, and a historical one further west in Cocos-Keeling (Bernardi et al., 2002; Leray et al., 2010; Salas et al., 2019). These clades are thought to have diverged allopatrically as a result of restricted gene flow between the two oceans at the Sunda Shelf during glacial maxima (Salas et al., 2020).

At the other end of the distribution in the Central Pacific, where most of the research on this complex has focused, previous studies have identified a number of species and isolated populations along the eastern boundary (Bernardi et al., 2002, 2003; Bernardi & Crane, 1999; Leray et al., 2010; Liggins et al., 2016; McCafferty et al., 2002; Ramon et al., 2008). In the subtropical Pacific islands of the Hawai'ian Archipelago and Johnston Atoll, *Dascyllus albisella* (Gill, 1862) is morphologically

distinguished by its pale coloration with black edged fins. Even within this relatively small distribution, both mitochondrial data (dloop and control region) and microsatellite data show population structure across its entire range (Leray et al., 2010; Ramon et al., 2008). To the southeast, *D. strasburgi* (Klausewitz, 1960) is an almost completely pale grey species with the most restricted range within the complex, found only in the Marquesas Islands. It seems to have been the first lineage of this complex to have diverged from the other species (McCafferty et al., 2002) despite its relative proximity to other regions within the complex' range. Genetic investigation on *Dascyllus auripinnis* (Randall & Randall, 2001), which was described based on morphometric differences including xanthic coloration, found that its distribution encompasses the Line and Phoenix Islands with a broad hybrid zone extending to the southwest to Fiji (Bernardi et al., 2002; Leray et al., 2010). It has been suggested that *D. auripinnis* likely diverged peripatrically and is now backcrossing with *D. trimaculatus* in the Central Pacific across this hybrid zone (Leray et al., 2010).

The far south-eastern edge population of *D. trimaculatus* in Mo'orea and around French Polynesia has been the site of much of the foundational research on the species and its evolutionary relationship with other species in the genus and with other *D. trimaculatus* populations (e.g. Bernardi et al., 2002; Bernardi & Crane, 1999; Leray et al., 2010; McCafferty et al., 2002). Based on mitochondrial and nuclear microsatellite data, French Polynesia too seems to host another peripheral diverged clade of the species complex (Bernardi et al., 2003; Leray et al., 2010). Here, one main clade includes only French Polynesian individuals, while in addition a small number of

individuals group with individuals collected in other regions of the Pacific in a sub-clade previously designated "OC3" (Bernardi et al. 2003, Leray et al. 2010).

Each set of studies focusing on these regions has added additional samples and methods, continually uncovering more structure and a better understanding of the evolutionary processes that have resulted in the modern-day *D. trimaculatus* complex. In this study, we shift our focus north-westward to the Ryukyu Archipelago, one of the northern-most coral reef systems and the range-limit to most Western Pacific coral reef species. One of the characteristics that make this regions interesting, is that despite its peripheral position in the sub-tropical latitudes of southern Japan, is that its location north of the Coral Triangle and its broad latitudinal span across temperature and environmental gradients not only helps to facilitate its rank as one of the most diverse marine regions but as a top-ranked center of endemism (Roberts et al., 2002). A key oceanographic feature driving the persistence of this unique system is the Kuroshio Current. This powerful current is formed when the west-bound Northern Equatorial Current (NEC) meets the Philippines and deflects northward becoming the Kuroshio. This current provides the Ryukyus, along with the rest of the East China Sea and the Pacific coast of southern Japan with warm tropical waters and propagules from the region known as center of tropical marine biodiversity (Carpenter & Springer, 2005). As it travels northward weaving through the Ryukyu Archipelago, the subsequent steady decline in sea surface temperature, light penetration, and aragonite saturation prevents the formation of hermatypic corals and coral reef habitat at the southern end of the main islands of Japan (Kleypas et al., 1999).

This current and archipelago provide a system to test the hypothesis that genetic structure can occur despite the existence of a strong mechanisms for gene flow, and allow us to examine whether *D. trimaculatus* is in fact less panmictic across its range than currently believed. From the Ryukyu region there is already research on scleractinian corals (Nakajima et al., 2010), zoantharians (Albinsky et al., 2018), crown-of-thorns sea stars (Yasuda et al., 2009), and fishes (Ackiss et al., 2018; Clark et al., 2021), with mixed results. For insight into how edge populations may be formed and their phylogeographic patterns within an otherwise widespread species, we compare this population to the French Polynesian population at the southern edge of the distribution. Previous genetic studies on *D. trimaculatus* biogeography used between 1-10 samples from the Ryukyus. In this study, we use the mitochondrial control region from 61 individuals sampled across the Ryukyu Archipelago along with 689 samples from across the *D. trimaculatus* complex range to evaluate the population structure of this unique edge population.

Materials and Methods

Tissue collection

Fin and/or gill tissue for the 52 individuals of *D. trimaculatus* was collected using pole spears or hand nets while scuba or free diving and preserved in 95% ethanol. Three sampling regions were chosen from across much of the extent of the Ryukyu Archipelago: Amami-Oshima (28.27°N,129.36°E), Okinawa (26.34°N,127.81°), and

Iriomote (24.33°N,123.82°E). Sampling was conducted at several sites around each island (Table 1). Other samples in this study are described in Supplementary Table 1. Sampling protocols were approved by Institutional Animal Care and Use Committee (IACUC) at the University of California Santa Cruz.

DNA extraction, amplification, and sequencing:

Total genomic DNA from the additional Ryukyu Archipelago samples was extracted using the Qiagen DNeasy Blood and Tissue kit. The mitochondrial control region (also called dloop) was amplified using forward primer CR-A (5'T TCC ACC TCT AAC TCC CAA AGC TAG 3' (Lee et al., 1995)) and a *D. trimaculatus* specific reverse primer DTPL (5'TTT GTT ACA GCA AAT TAT TTAT3' (Bernardi et al., 2002)). Polymerase chain reactions (PCR) were performed in 20µl reactions which consisted of 1µl DNA, 1µl of forward and 1 µl reverse primer (10p), 1µl MgCl, 2µl Coral Red Buffer dye and 4 µl of autoclaved MilliQ water. The PCR amplification profile began with denaturation at 94°C for 45 s, annealing at 50°C for 45 s, extension at 72°C for 1 min, and repeated for 35 cycles. DNA was cleaned using STAR following manufacturers protocol and sanger sequenced (FASMAC, Kanagawa Japan). Sequences were aligned and edited using Geneious (version 10.2.6).

Analysis

Phylogenetic relationships were evaluated using aligned sequences for *Dascyllus trimaculatus* and a single *D. reticulatus* sequence as an outgroup following

Leray et al. (2010). Tree reconstructions were made using PAUP (v4.0) (Swofford), with distance parameters (Kimura 2) used to reconstruct Neighbor-Joining trees (Nei). Metrics of population structure (Fixation indexes, Fst) were estimated using Genodive (v.3.0.6) (Meirmans) after creating a genepop input file using PGD Spider (Excoffier), and DNAsp (v. 6.0) (Rozas and Rozas) using the PAUP nexus files as input files.

To determine whether the population distribution pattern was the same for all groups represented in clade V, the proportion of Ryukyu samples that were in each clade were calculated. Those proportions were set to be the expected values for each of the other groups of samples (Philippines, Indonesia, Fiji+Wallis), and a chi-square test was performed.

Results

Phylogenetic results:

The mtDNA data partitioned the *Dascyllus trimaculatus* complex into five major clades that corresponded to *D. albisella* in Hawaii, *D. strasburgi* in Marquesas, *D. trimaculatus* in the Indian Ocean, *D. trimaculatus* in the Pacific, and *D. trimaculatus* in French Polynesia (Figure 2). Samples from a single locality tended to group within one major clade, except for French Polynesian samples that clustered into two main clades. One clade was comprised only of French Polynesian individuals, while the other clade included both French Polynesian samples, as well as other Pacific individuals in Pacific sub-clade I (previously OC3 clade (Bernardi et al., 2003; Leray et al., 2010)).

The Pacific clade was itself divided into seven of these subclades (I-VII) (Figure 3). Data showed that of the 61 Ryukyu (Japan) samples 54 belonged to subclade V, three were in subclade II and four in the previously mentioned subclade I with French Polynesian samples. This is a significantly uneven distribution (88% of individuals belong to subclade V; chi-square, p value=$5.9x10^{-55}$) and a pattern that is significantly different from the distribution of individuals that belong to other populations that are present in these clades (Fiji+Wallis Island: n=20, chi square test, p value=$1x10^{-27}$; Philippines: n=13, chi square test, p value=$9.6x10^{-12}$; Indonesia: n=9, chi square test, p value=$8.3x10^{-16}$)

Genetic diversity and gene flow

Within the Japanese population, the average pairwise nucleotide difference (Pi) was 0.0158 (1.58%) or 3.608 nucleotides (k). In the French Polynesian population, the average pairwise nucleotide difference and the average pairwise number of differing nucleotides were Pi=0.0286 and k=9.76, respectively, and when analyzed again without the individuals in clade I, Pi=0.02451 and k=8.359.

Gene flow and estimated migration rates between Japanese samples and the rest of the Pacific Rim was low (Fst=0.05685; Nm=8.29). The other edge population, French Polynesia, showed even more limited gene flow with the Pacific Rim (Fst=0.6976, Nm=0.22).

Discussion

Genetic and genomic markers have been used to better understand the phylogeography of the three-spot damselfish revealing important insights into the evolutionary relationships and mechanisms of speciation within the *Dascyllus* complex. Most of these studies have focused on the eastern portion of the species distribution in the central Pacific (Bernardi et al., 2001, 2003; Bernardi & Crane, 1999; Leray et al., 2010; Liggins et al., 2016; Ramon et al., 2008), as well as in the Indian Ocean and on Indo-Pacific relationships (Salas et al., 2020, 2019).

Data presented here are consistent with previously published results, where *Dascyllus albisella* in Hawaii and *D. strasburgi* in the Marquesas Islands each formed monophyletic assemblages, thus matching currently accepted taxonomic rankings (Bernardi et al., 2001, 2003; Leray et al., 2010). Figure 2. Individuals identified as *D. trimaculatus* group in very distinct clades, for example Indian Ocean, Pacific Ocean and French Polynesian samples cluster in different groups, indicating that taxonomic revision will be needed for these putative taxa once additional nuclear markers are used to confirm results based on mitochondrial markers.

Here, we present phylogeographic results using the most robust set of *Dascyllus trimaculatus* samples collected in the Ryukyu Archipelago to bring focus to this north-west distribution edge. We compare with and gain insight from another more well-studied edge population in French Polynesia.

The Pacific Rim clade itself shows interesting divisions and is composed of at least seven distinct clades (I-VII, Figure 2, 3). Most individuals collected in the Ryukyu Archipelago group in subclade V with only few individuals each in subclades I and II. Gene flow (Fst=0.57) and migration rate (Nm=8.29) indicated that despite its proximity to the Coral Triangle and position along the Kuroshio Current, the Ryukyu population can be characterized as its own, fairly closed population. Phylogenetic analyses showed that the Ryukyu population also has individuals which may have arrived from elsewhere in the Pacific and conversely the Ryukyu dominant clade is also found in other regions such as in Indonesia, the Philippines, and as far away as Fiji and Wallis Island. Subclade V is also distinct among the Pacific Rim subclades in the shallow depth of its basal branch, potentially suggesting both a recent colonization of the region (lack of lineage sorting), and the immigration of ancient polymorphisms (basal branches). Indeed, non-Japanese clade V individuals tended to cluster in the more basal branches of the clade, which may point to the ancestral source of the original colonizing population. Nevertheless, it is apparent that there is gene flow into the Ryukyu populations as shown by the individuals from clade I and II found in the archipelago. The recent nature of the Japanese population may be the signature of a more dynamic edge population at the margins of coral reefs ecosystems.

Although the Ryukyus are home to prominent coral reef habitats, their location in the subtropics at ~24-31N° latitude means that there is relatively large seasonal fluctuation in ocean temperature (16.4°C-33.3°C) than in lower latitude regions

(25.8°C-30.4°C) (Iguchi & Hongo, 2018; Kleypas et al., 1999). This high environmental variation makes for a physiologically more stressful place for organisms with restricted mobility to live in. Over longer timescales, the coral reefs of the Ryukyus may rely on the delivery of warm equatorial water (and propagules) by the Kuroshio Current, the strength and path of which is determined by seasonal factors such as wind, typhoons, and temperatures, decadal phenomena that may accentuate the seasonal factors and more long-term factors such as glaciation periods, which change sea levels and therefore flow paths, thus also influencing all the previously listed above factors (Chen, 2013). It does seem plausible that such fluctuations result in occasional die offs of entire guilds of the most susceptible fishes that are restricted to shallow waters, as has been seen in zooxanthellate scleractinian corals in both abnormally cold winters (e.g.(Suzuki et al., 2013) and warm summers (Masucci et al., 2019; Omori, 2011) in the region. If this is the case, re-colonization of this edge population is likely to be frequent and when observed, recent.

South across the equator and at the eastern edge of the *D. trimaculatus* distribution in French Polynesia, we can see a mirror image of the Ryukyu population distribution (Fig 3). It is composed of a main clade (French Polynesian clade) with additional individuals from another deep branched Pacific Rim subclade I. Like the Ryukyu population, *D. trimaculatus* in French Polynesia has very little connection with other populations with an estimated 0.22 migrants per generation and Fst=0.6976. Its basal

branch is far more pronounced, however, and unlike the Ryukyus, we interpret this as a deep and old divergence of this relatively closed population.

French Polynesia also differs from the Ryukyus in other significant ways. These islands sit in lower tropical latitudes (17.5° S as opposed to the Ryukyus at 31°N) and experience a much more stable environment with seasonal sea surface temperature ranges that vary by less than 5°C compared to almost 17°C ranges in the Ryukyus (Kleypas et al., 1999). Unlike the Kuroshio's effect on the Ryukyus, Leray et al (2010) suggested that the currents in the French Polynesia region have had an isolating effect on *D. trimaculatus*.

The comparison between our current results of the Ryukyus population and the more well understood French Polynesian population offer an interesting perspective on processes that may be at play in each region. While they are both clearly isolated populations, we are likely looking at different stages of diverging populations. French Polynesia is far more well defined, with clades on deep branches, a fact which is further supported by microsatellite data (Leray et al., 2010). Our current study is the only research to date which has looked specifically at *D. trimaculatus* in southern Japan, and our mtDNA data suggest a comparatively much younger population perhaps in the early stages of coalescence. The differences in the respective environments also raises the question of whether the Ryukyus, as a marginal environment, experience "regular" local extirpations and colonization events, explaining the short basal branch of clade V

and supporting the general scientific view that edge populations are often far more vulnerable to extinction events.

Acknowledgements

Associate Prof. Tohru Naruse (University of the Ryukyus) in Iriomote and Dr. Takuma Fujii (Kagoshima University) in Amami-Oshima are thanked for logistical help. Permits and permission obtained for all sampling in Japan. This work was supported by the East Asia and Pacific Summer Institute (EAPSI) Fellowship funded by JSPS and NSF to MBR.

References

Ackiss, A. S., Bird, C. E., Akita, Y., Santos, M. D., Tachihara, K., & Carpenter, K. E. (2018). Genetic patterns in peripheral marine populations of the fusilier fish Caesio cuning within the Kuroshio Current. *Ecology and Evolution*, *8*(23), 11875–11886. https://doi.org/10.1002/ece3.4644

Albinsky, D., Wham, D., Shinzato, N., & Reimer, J. D. (2018). Population Connectivity in the Common Reef Zoantharian Zoanthus sansibaricus (Anthozoa: Hexacorallia) in Southern Japan. *Zoological Science*, *35*(4), 321. https://doi.org/10.2108/zs180007

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709. https://doi.org/10.1038/nrg2844

Bernardi, G., & Crane, N. L. (1999). Molecular phylogeny of the humbug damselfishes inferred from mtDNA sequences. *Journal of Fish Biology*, *54*(6), 1210–1217. https://doi.org/10.1111/j.1095-8649.1999.tb02049.x

Bernardi, G., Holbrook, S. J., & Schmitt, R. J. (2001). Gene Flow at three spatial scales in a coral reef Fish, the three-spot dascyllus, Dascyllus trimaculatus. *Marine Biology*, *138*, 457–465.

Bernardi, G., Holbrook, S. J., Schmitt, R. J., & Crane, N. L. (2003). Genetic evidence for two distinct clades in a French Polynesian population of the coral reef three-spot damselfish Dascyllus trimaculatus. *Marine Biology*, *143*(3), 485–490. https://doi.org/10.1007/s00227-003-1091-y

Bernardi, G., Holbrook, S. J., Schmitt, R. J., Crane, N. L., & DeMartini, E. (2002). Species boundaries, populations and colour morphs in the coral reef three-spot damselfish (Dascyllus trimaculatus) species complex. *Proceedings of the Royal Society B: Biological Sciences*, *269*(1491), 599–605. https://doi.org/10.1098/rspb.2001.1922

Carpenter, K. E., & Springer, V. G. (2005). The center of the center of marine shore fish biodiversity: The Philippine Islands. *Environmental Biology of Fishes*, *72*(4), 467–480. https://doi.org/10.1007/s10641-004-3154-4

Chen, F. (2013). *The Kuroshio Power Plant* (Vol. 15). Springer International Publishing. https://doi.org/10.1007/978-3-319-00822-6

Clark, R. D., Aardema, M. L., Andolfatto, P., Barber, P. H., Hattori, A., Hoey, J. A., Montes, H. R., & Pinsky, M. L. (2021). Genomic signatures of spatially divergent selection at clownfish range margins. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1952), 20210407. https://doi.org/10.1098/rspb.2021.0407

Iguchi, A., & Hongo, C. (Eds.). (2018). *Coral Reef Studies of Japan* (Vol. 13). Springer Singapore. https://doi.org/10.1007/978-981-10-6473-9

Kleypas, J. A., Mcmanus, J. W., & Meñez, L. A. B. (1999). Environmental Limits to Coral Reef Development: Where Do We Draw the Line? *American Zoologist*, *39*(1), 146–159. https://doi.org/10.1093/icb/39.1.146

Lee, W.-J., Conroy, J., Howell, W. H., & Kocher, ThomasD. (1995). Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, *41*(1). https://doi.org/10.1007/BF00174041

Leray, M., Beldade, R., Holbrook, S. J., Schmitt, R. J., Planes, S., & Bernardi, G. (2010). Allopatric Divergence and Speciation in Coral Reef Fish: The Three-Spot Dascyllus, Dascyllus Trimaculatus, Species Complex. *Evolution*, *64*(5), 1218–1230. https://doi.org/10.1111/j.1558-5646.2009.00917.x

Liggins, L., Treml, E. A., Possingham, H. P., & Riginos, C. (2016). Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. *Journal of Biogeography*, *43*(2), 256–267. https://doi.org/10.1111/jbi.12647

Masucci, G. D., Biondi, P., Negro, E., & Reimer, J. D. (2019). After the long summer: Death and survival of coral communities in the shallow waters of Kume Island, from the Ryukyu Archipelago. *Regional Studies in Marine Science*, *28*, 100578. https://doi.org/10.1016/j.rsma.2019.100578

McCafferty, S., Bermingham, E., Quenouille, B., Planes, S., Hoelzer, G., & Asoh, K. (2002). Historical biogeography and molecular systematics of the Indo-Pacific genus Dascyllus (Teleostei: Pomacentridae). *Molecular Ecology*, *11*(8), 1377–1392. https://doi.org/10.1046/j.1365-294X.2002.01533.x

Nakajima, Y., Nishikawa, A., Iguchi, A., & Sakai, K. (2010). Gene Flow and Genetic Diversity of a Broadcast-Spawning Coral in Northern Peripheral Populations. *PLoS ONE*, *5*(6), e11149. https://doi.org/10.1371/journal.pone.0011149

Omori, M. (2011). Degradation and restoration of coral reefs: Experience in Okinawa, Japan. *Marine Biology Research*, *7*(1), 3–12. https://doi.org/10.1080/17451001003642317

Ramon, M. L., Nelson, P. A., De Martini, E., Walsh, W. J., & Bernardi, G. (2008). Phylogeography, historical demography, and the role of post-settlement ecology in two Hawaiian damselfish species. *Marine Biology*, *153*(6), 1207–1217. https://doi.org/10.1007/s00227-007-0894-7

Randall, J. E., & Randall, H. A. (2001). Dascyllus auripinnis, a New Pomacentrid Fish from Atolls of the Central Pacific Ocean. *Zoological Studies*, 7.

Roberts, C. M., McClean, C. J., Veron, J. E. N., Hawkins, J. P., Allen, G. R., McAllister, D. E., Mittermeier, C. G., Schueler, F. W., Spalding, M., Wells, F., Vynne, C., & Werner, T. B. (2002). Marine Biodiversity Hotspots and Conservation Priorities for Tropical Reefs. *Science*, *295*(5558), 1280–1284. https://doi.org/10.1126/science.1067728

Salas, E. M., Hobbs, J. A., Bernal, M. A., Simison, W. B., Berumen, M. L., Bernardi, G., & Rocha, L. A. (2020). Distinct patterns of hybridization across a suture zone in a coral reef fish (*Dascyllus trimaculatus*). *Ecology and Evolution*, *10*(6), 2813–2837. https://doi.org/10.1002/ece3.6068

Salas, Eva. M., Bernardi, Giacomo., Berumen, Michael. L., Gaither, Michelle. R., & Rocha, Luiz. A. (2019). RADseq analyses reveal concordant Indian Ocean biogeographic and phylogeographic boundaries in the reef fish *Dascyllus trimaculatus*. *Royal Society Open Science*, *6*(5), 172413. https://doi.org/10.1098/rsos.172413

Suzuki, G., Yatsuya, K., & Muko, S. (2013). Bleaching of tabular Acropora corals during the winter season in a high-latitude community (Nagasaki, Japan). *Galaxea, Journal of Coral Reef Studies*, *15*(2), 43–44. https://doi.org/10.3755/galaxea.15.43

Yasuda, N., Nagai, S., Hamaguchi, M., Okaji, K., Gérard, K., & Nadaoka, K. (2009). Gene flow of Acanthaster planci (L.) in relation to ocean currents revealed by microsatellite analysis. *Molecular Ecology*, *18*(8), 1574–1590. https://doi.org/10.1111/j.1365-294X.2009.04133.x

Ch1 - Tables and Figures

Table 1. All *Dascyllus trimaculatus* samples collected in the Ryukyu Archipelago used in this study are listed here with location, sample ID and corresponding sample ID listed in Supplementary Figure.

Table 1.

		sample ID	mtDNA ID	site	coordinates
Ryukyu Archipelago, Japan (total n=61)	Amami-Oshima (n=18)	DTR_AMA-071701	AMA-96	Arira	28.444819, 129.527271
		DTR_AMA-071702	AMA-97	Arira	28.444819, 129.527271
		DTR_AMA-071703	AMA-98	Shirahama	28.196805, 129.271029
		DTR_AMA-071704	AMA-99	Shirahama	28.196805, 129.271029
		DTR_AMA-071705	AMA-100	Sesui	28.134974, 129.330211
		DTR_AMA-071706	AMA-101	Sesui	28.134974, 129.330211
		DTR_AMA-071709	AMA-104	Sesui	28.134974, 129.330211
		DTR_AMA-071710	AMA-105	Sesui	28.134974, 129.330211
		DTR_AMA-071711	AMA-106	Sesui	28.134974, 129.330211
		DTR_AMA-071712	AMA-107	Sesui	28.134974, 129.330211
		DTR_AMA-071713	AMA-108	Sesui	28.134974, 129.330211
		DTR_AMA-071714	AMA-109	Sesui	28.134974, 129.330211
		DTR_AMA-071715	AMA-110	Sesui	28.134974, 129.330211
		DTR_AMA-071716	AMA-111	Sesui	28.134974, 129.330211
		DTR_AMA-071717	AMA-112	Sesui	28.134974, 129.330211
		DTR_AMA-071718	AMA-113	Sesui	28.134974, 129.330211
		DTR_AMA-071719	AMA-114	Sesui	28.134974, 129.330211
		DTR_AMA-071720	AMA-115	Shidokan	28.328324, 129.300228
	Okinawa (n=34)	DTR_OHED_071701	OHED-134	Hedo Port	26.854556, 128.246460
		DTR_OHED_071702	OHED-135	Hedo Port	26.854556, 128.246460
		DTR_OHED_071702b	OHED-135b	Hedo Port	26.854556, 128.246460
		DTR_OHED_071703	OHED-136	Hedo Port	26.854556, 128.246460
		DTR_OHED_071704	OHED-137	Hedo Port	26.854556, 128.246460
		DTR_OHED_071705	OHED-138	Hedo Port	26.854556, 128.246460
		DTR_OHED_071706	OHED-139	Hedo Port	26.854556, 128.246460
		DTR_OHED_071707	OHED-140	Hedo Port	26.854556, 128.246460
		DTR_OHED_071708	OHED-189	Hedo Port	26.854556, 128.246460
		DTR_OHED_071709	OHED-190	Hedo Port	26.854556, 128.246460
		DTR_OHED_071710	OHED-191	Hedo Port	26.854556, 128.246460
		DTR_OHED_071711	OHED-192	Hedo Port	26.854556, 128.246460
		DTR_OHED_071712	OHED-193	Hedo Port	26.854556, 128.246460
		DTR_OHED_071715	OHED-196	Hedo Port	26.854556, 128.246460
		DTR_OHED_071716	OHED-197	Hedo Port	26.854556, 128.246460
		DTR_OHED_071717	OHED-198	Hedo Port	26.854556, 128.246460
		DTR_OHED_071718	OHED-199	Hedo Port	26.854556, 128.246460
		DTR_OMIA_071702	OMIA-201	Miyagi-jima	26.364559, 127.997884
		DTR_OMIA_071703	OMIA-202	Miyagi-jima	26.364559, 127.997884
		DTR_OOUR_071701	OOUR-203	Oura Bay	26.536729, 128.079806
		DTR_OOUR_071702	OOUR-204	Oura Bay	26.536729, 128.079806
		DTR_OZPA_071702	ZPA-206	Zanpa	26.441628, 127.712330
		DTR_OZPA_071703	ZPA-207	Zanpa	26.441628, 127.712330
		DTR_OZPA_071704	ZPA-208	Zanpa	26.441628, 127.712330
		DTR_OZPA_071705	ZPA-209	Zanpa	26.441628, 127.712330
			OKI1	Okinawa	
			OKI2	Okinawa	
			OKI3	Okinawa	
			OKI5	Okinawa	
			OKI6	Okinawa	
			OKI8	Okinawa	
			OKI7	Okinawa	
			AKA1	Aka-jima (Okinawa)	
			AKA2	Aka-jima (Okinawa)	
	Iriomote-jima (n=9)	DTR_IRI__071701	IRI-117	Sotobanari	24.361374, 123.706348
		DTR_IRI__071702	IRI-118	near Hoshizuna and Maruma	24.434927, 123.790206
		DTR_IRI__071703	IRI-119	reef off Hatoma	24.452296,123.815298
		DTR_IRI__071704	IRI-120	reef off Hatoma	24.452296,123.815298
		DTR_IRI__071705	IRI-121	reef off Hatoma	24.452296,123.815298
		DTR_IRI__071707	IRI-123	reef off Hatoma	24.453646, 123.810382
		DTR_IRI__071708	IRI-124	reef off Hatoma	24.453646, 123.810382
		DTR_IRI__071709	IRI-125	reef off Hatoma	24.453646, 123.810382
		DTR_IRI__071712	IRI-128	Maruma Beach	24.424088,123.802420

Figure 1. Map of distributions of the *Dascyllus trimaculatus* species complex with illustrated example of a specimen from each distinct group along the top. Background color of each fish match color of distribution shown in map. Red square shows location of the Ryukyu archipelago.

Figure 1.

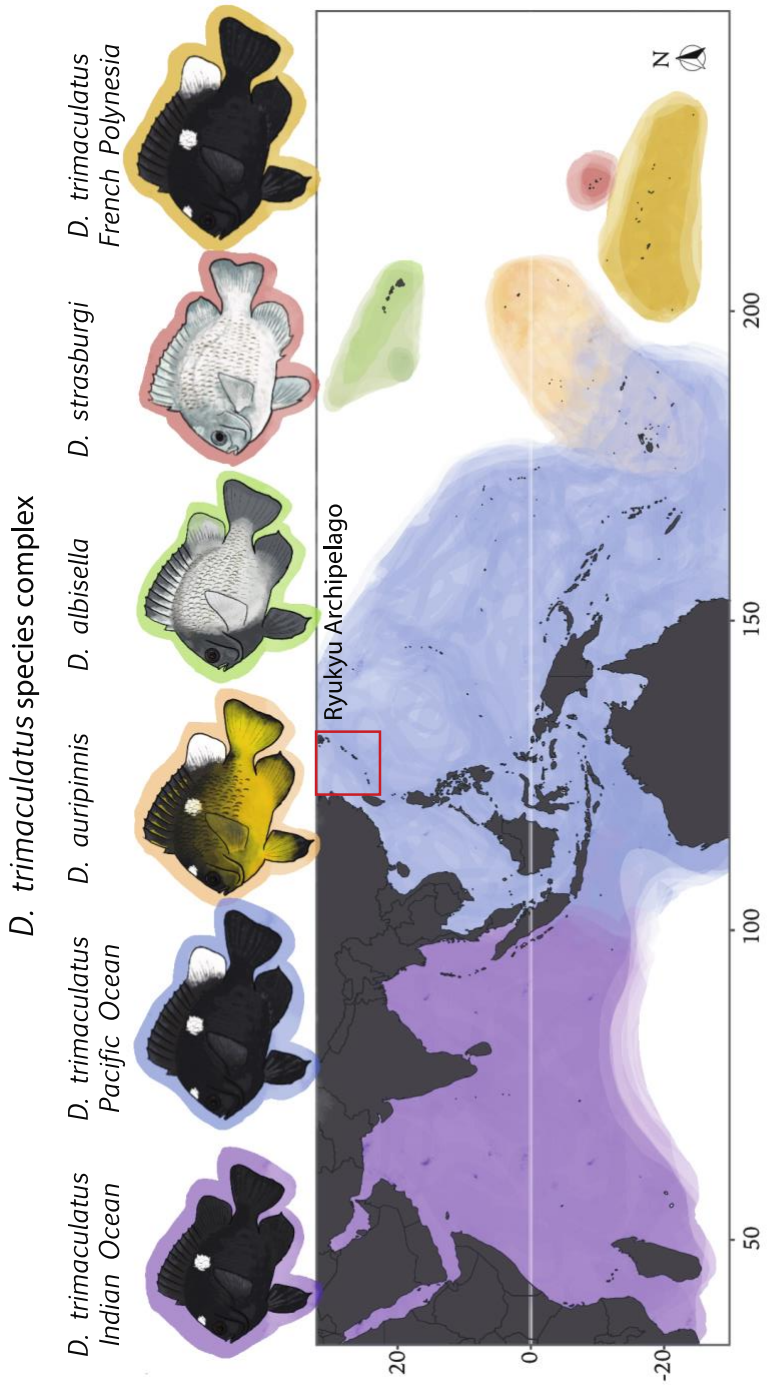


Figure 2. Neighbor joining-tree using all Ryukyu, Indian Ocean, Pacific Rim, Marquesas samples, a random subset of French Polynesian and Hawai'ian samples and

one *D. reticulatus* sample for the out group. Branch colors correspond to major clades of the *Dascyllus trimaculatus* species complex labelled in outer circle. Roman numerals (I-VI) number each of the clades within the Pacific Rim clade. Colors around the edge of the tree correspond to the location from which that sample was collected described in the figure legend.

Figure 2

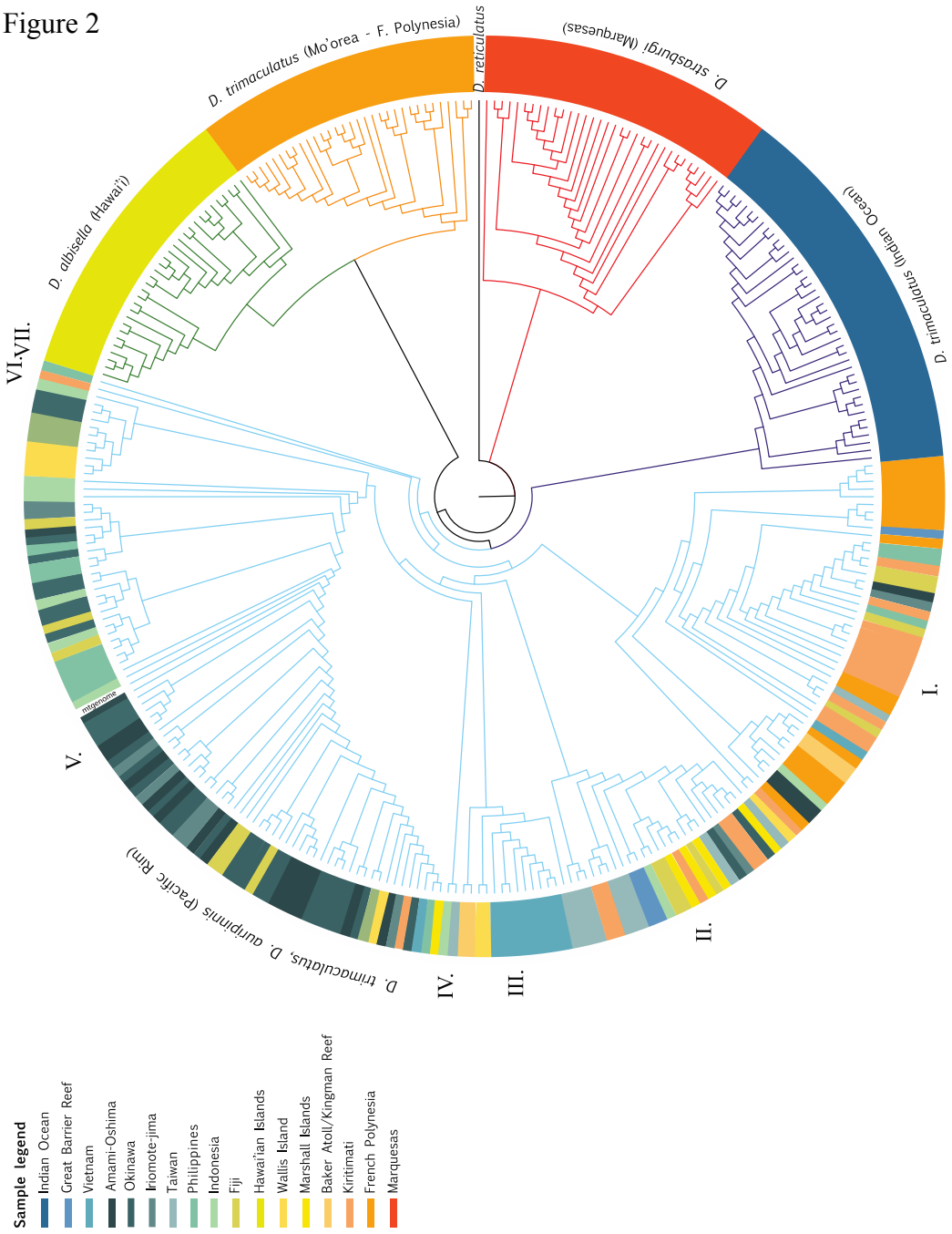
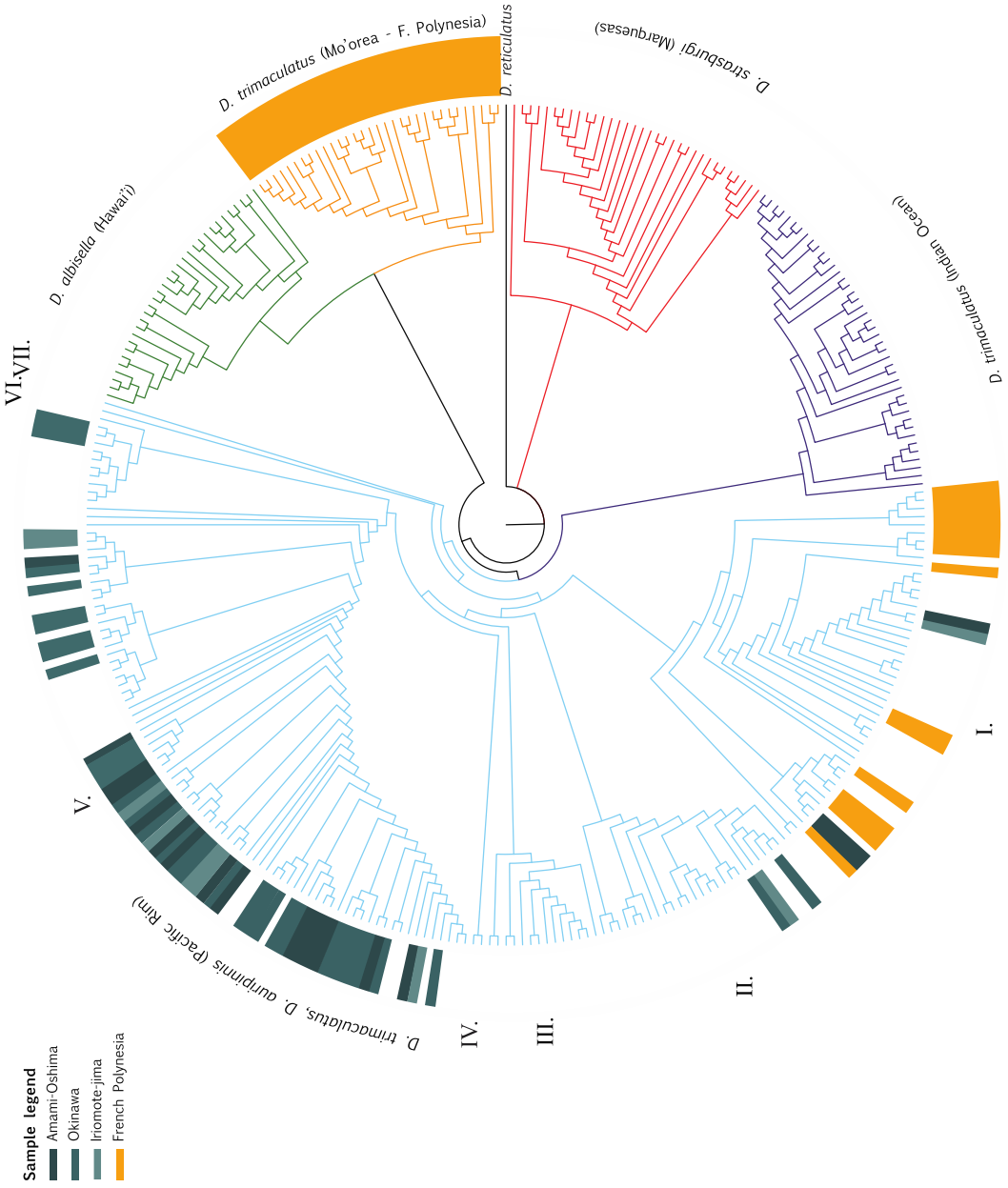

D. reticulatus
D. strasburgi (Marquesas)
D. trimaculatus (Indian Ocean)
D. trimaculatus (Mo'orea - F. Polynesia)
D. albisella (Hawai'i)
D. trimaculatus, D. auripinnis (Pacific Rim)
I.
II.
III.
IV.
V.
VI-VII.
outgroup
Sample legend
Indian Ocean
Great Barrier Reef
Vietnam
Amami-Oshima
Okinawa
Iriomote-jima
Taiwan
Philippines
Indonesia
Fiji
Hawai'ian Islands
Wallis Island
Marshall Islands
Baker Atoll/Kingman Reef
Kiritimati
French Polynesia
Marquesas

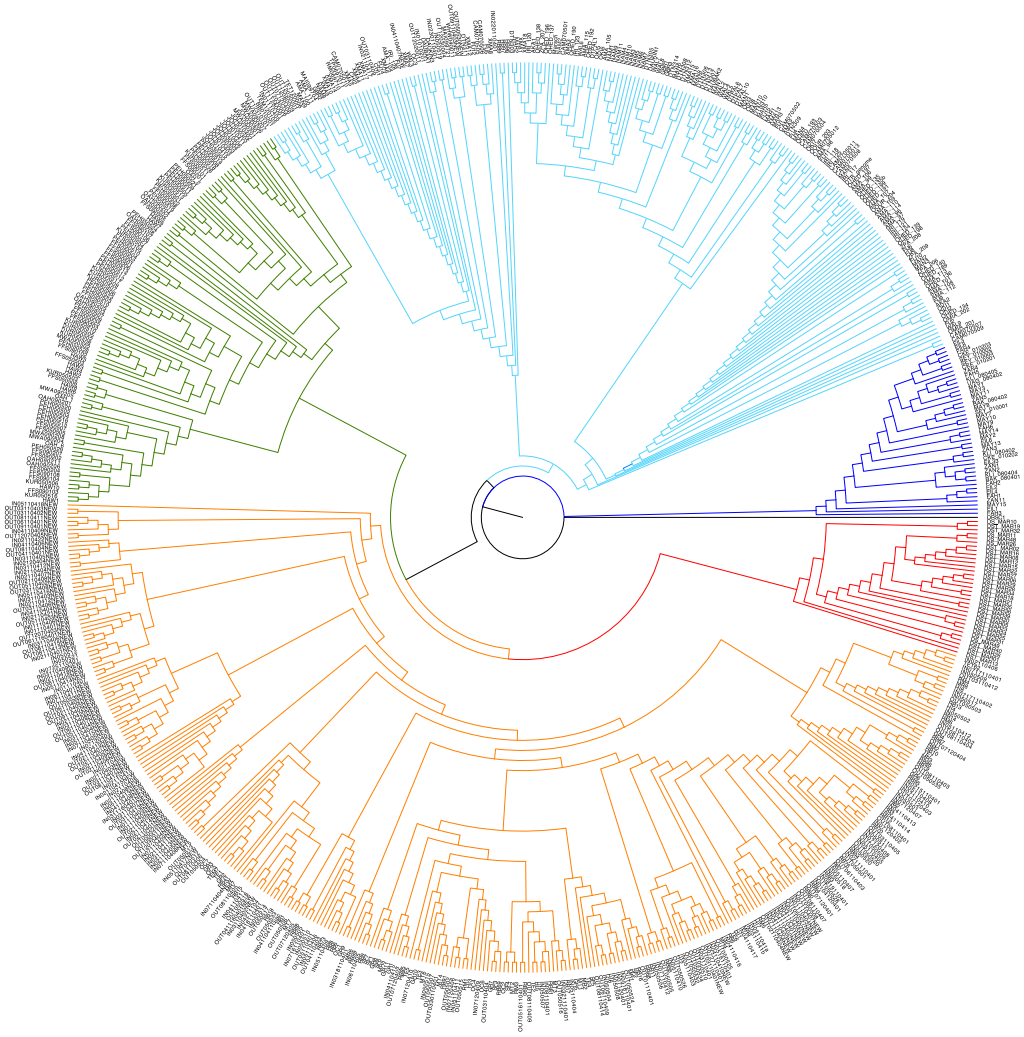
Figure 3. Neighbor joining-tree using all Ryukyu, Indian Ocean, Pacific Rim, Marquesas samples, a random subset of French Polynesian and Hawai'ian samples and one *D. reticulatus* sample for the out group. Branch colors correspond to major clades of the *Dascyllus trimaculatus* species complex labelled in outer circle. Roman numerals (I-VI) number each of the clades within the Pacific Rim clade. Colors around the edge of the tree correspond to the location from which that sample was collected described in the figure legen*d.* This tree highlights samples from the Ryukyus and French Polynesian samples only. It is otherwise identical to Figure 2 of this paper.

Figure 3.


D. trimaculatus (Mo'orea - F. Polynesia)
D. reticulatus
D. strasburgi (Marquesas)
D. trimaculatus (Indian Ocean)
D. albisella (Hawai'i)
VI.
VII.
V.
IV.
III.
II.
I.
D. trimaculatus, D. aurigninis (Pacific Rim)
Sample legend
Amami-Oshima
Okinawa
Iriomote-jima
French Polynesia

Supplementary Figure 1. Neighbor joining-tree using all 689 samples of *Dascyllus trimaculatus* and one *D. reticulatus* sample for the out group. Branch colors correspond to colors used in previous figures identifying major clades of the *Dascyllus trimaculatus* species complex.

Figure S1.



Chapter 2

Chromosome-level genome of the three-spot damselfish, *Dascyllus trimaculatus*

May B. Roberts[1], Darrin T. Schultz[2,3,4], Remy Gatins[5],
Merly Escalona[4], Giacomo Bernardi[1]

1. Department of Ecology and Evolutionary Biology, 115 McAllister Rd, University of California Santa Cruz, CA, USA
2. Department of Molecular Evolution and Development, University of Vienna, Vienna, 1010, Austria
3. Monterey Bay Aquarium Research Institute, Moss Landing, California 95039, USA
4. Department of Biomolecular Engineering and Bioinformatics, University of California, Santa Cruz, California 95064, USA
5. Department of Marine Sciences, Northeastern University, MA, USA

Abstract

Damselfishes (Family: Pomacentridae) are a group of ecologically important, primarily coral reef fishes that include over 400 species. Damselfishes have been used as model organisms to study recruitment (anemonefishes), the effects of ocean acidification (spiny damselfish), population structure and speciation (*Dascyllus*). The genus *Dascyllus* includes a group of small bodied species, and a complex of relatively larger bodied species, the *Dascyllus trimaculatus* species complex that comprises several species including *D. trimaculatus* itself. The three-spot damselfish, *D. trimaculatus* is a widespread and common coral reef fish species found across the tropical Indo-Pacific. Here we present the first genome assembly of this species. This assembly contains 910 Mb, 90% of the bases are in 24 chromosome-scale scaffolds, and the BUSCO score of the assembly is 97.9%. Our findings confirm previous reports of a karyotype of 2n = 47 in *D. trimaculatus* in which one parent contributes 24 chromosomes and the other 23. We find evidence that this karyotype is the result of a heterozygous Robertsonian fusion. We also find that the *D. trimaculatus* chromosomes are each homologous with single chromosomes of the closely related clownfish species, *Amphiprion percula*. This assembly will be a valuable resource in the population genomics and

conservation of Damselfishes, and continued studies of the karyotypic diversity in this clade.

Introduction

Damselfishes (Pomacentridae) are a group of small-bodied species found across all coral reef regions and most temperate marine systems where they are often the most visibly abundant fishes on the reef (1–4). This family includes more than 400 species that, despite their small size (max 30cm), play important ecological roles (2,5). Within this large family, the genus *Dascyllus* comprises eleven species, four of which, make up the *Dascyllus trimaculatus* species complex. This species complex includes three described species with restricted geographic ranges, *D. albisella* in the Hawaiian Islands, *D. strasburgi* in the Marquesas Islands, and *D. auripinnis* in the Line Islands. In contrast, *D. trimaculatus* has the broadest range, extending from the Red Sea, where it was first described (Ruppell, 1829), across the tropical and subtropical Indo-Pacific (Figure 1).

Three-spot damselfish is an abundant and common species, that exhibit a typical bipartite life history, with a site attached adult phase, where mate pairs lay, fertilize, and care for demersal eggs, followed by a pelagic larval phase. Larvae hatch after ~6 days and feed in the water column on zooplankton where their pelagic larval duration (PLD) lasts 23-30 days until they recruit back to the reef (6,7). Larvae settle primarily into anemones for protection often sharing this shelter with different species of the popular

anemonefish (in Hawai'i, where anemonefishes and anemones are absent, *D. albisella* recruits to branching coral). As subadults, they leave the anemone and live nearby in small to large groups.

There has also been considerable effort in understanding the chromosomes architecture and variation of *Dascyllus* and other damselfishes. Chromosome number varies between species of *Dascyllus* as well as within species (8–10) giving insight into chromosomal drivers of evolution (11–13) and how this variation is manifested ecologically (11,14). As we shift into the age of genomic natural history where genomic tools offer vastly more detail and statistical power, a reference genome will aid in further refining our understanding of wildlife biology (15). There are currently 14 Pomacentrid reference genomes, five of which are publicly available through the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) (*A. ocellaris*, (16), *Acanthochromis polyacanthus*, (17); *Amphiprion percula*, (18); *Amphiprion ocellaris* (19); *Acanthochromis polyacanthus*, Lehmann in review), and another nine from a single study (Marcionetti et al. 2019). Of these, only one is of species other than genus *Amphiprion* and only three of those listed above (*A. ocellaris*, *A. percula, A. polyacanthus*) are chromosome-scale genomes. Of the Pomacentrid chromosome-scale genomes all had 2n=48, with genome sizes ranging

between 863Mb-956Mb. The two published genomes, *A. ocellaris*, (Ryu et al. 2022) and *A. percula* were highly complete with published BUSCO values of 97.01% and 97.2% respectively. Chromosome-scale genomes provide a more complete sequence and locations of genes and allow for research into how chromosome architecture influences ecology, population dynamics, and adaptive evolution. Here, we present the first genome assembly within the genus *Dascyllus* and add to the short, but growing list of Pomacentrid chromosome-scale genomes.

Materials and Methods

Biological Materials:

The *D. trimaculatus* individual used for this genome assembly was ordered from an online pet fish supplier (liveaquaria.com), sourced from the West Pacific Rim population (20). It was euthanized following an approved IACUC protocol animal use. Liver, muscle, gill, and brain tissue were harvested from the right side of the individual and each placed in separate, pre-weighed Covaris cryogenic vials, flash frozen in liquid nitrogen, and stored at -80°C until further processing. The remaining intact left-side of the specimen is stored in -80°C at University of California Santa Cruz.

Nucleic acid library preparation and sequencing:

Whole-genome shotgun library preparation:

DNA was extracted from 13 mg of muscle tissue using a DNeasy Blood and Tissue kit (Qiagen), quantified using Qubit dsDNA HS Assay kit (Thermo Fisher Scientific) and Qubit 4.0 Fluorometer, then assayed with 1.0% agarose gel electrophoresis to determine molecular weight. DNA was sheared for 26 cycles of shearing (15s on, 30s chilling) using a Bioruptor sonicator (Diagenode), then size selected using SPRI beads (Beckman) to select for fragments between 200 bp to 500 bp.

The NEBNext UltraII DNA Library Prep Kit for Illumina (New England Bio Labs) was used according to manufacturer's protocol except that KAPA Hot Mix Ready Start Master Mix (Roche Diagnostics) was used for library amplification instead of NEB Q5 Master Mix. Paired-end sequencing was done at the University of California Davis Genome Center on a HiSeq4000 sequencer on a 2x150PE cycle.

Chicago library preparation

High molecular weight (HMW) DNA was isolated from the *Dascyllus trimaculatus* individual by lysing gill tissue in low-EDTA TE buffer (21), then purifying with a chloroform, phenol:chloroform, chloroform, ethanol

precipitation protocol (22). The quality of the HMW DNA was assayed with 1.0% agarose gel electrophoresis. This DNA was used in the preparation of the Chicago, Hi-C, and for Oxford Nanopore Technologies sequencing libraries.

From this DNA, three Chicago libraries were prepared using a published method (23), each using a different restriction enzyme: one with DpnII cutting at GATC sites, one with MluCI cutting at AATT sites, and one with FatI cutting at CATG sites. These libraries were sequenced on a 2x150PE cycle at Fulgent Genetics on a HiSeq400 sequencer.

Hi-C library preparation

Two Hi-C libraries were generated from approximately 100 ng of LN_2-flash-frozen muscle. The libraries were constructed using a published protocol (24). One library was constructed using the enzyme DpnII, and the other library was constructed with the enzyme MluCI.

Oxford Nanopore library (ONT):

Next, 1500 ng of the HMW DNA prepared for Chicago libraries was also used to prepare two ONT WGS libraries with the SQK-LSK109 modified protocol "versionGDE_9063_v109_revT_14Aug2019". The DNA repair steps at

20°C and 65°C were carried out for 20 minutes each, instead of 5 minutes each. We ran each of the resulting libraries on two separate MinION flow cells (FLO-MIN106), each for 72 hours. Raw fast5 files from the two MinION runs were basecalled using Guppy(25) v3.3.

A summary of sequencing information for the various libraries can be found in Supplementary Table 1.

Genome Assembly:

All programs and versions used for the assembly are listed in Table 1. Sequencing adapters were removed from the Illumina whole-genome shotgun (WGS) reads with Trimmomatic (26) v0.39 with parameters: 'ILLUMINACLIP: all_seqs.fa:2:30:10:8:TRUE SLIDINGWINDOW:4:20 MINLEN:50'. We used jellyfish (26) v2.2.10 to make a k-21 k-mer count vs abundance histogram and used the histogram with Genome Scope (27) v2.0 to estimate *D. trimaculatus* genome size, heterozygosity, and repeat content. MaSuRCA (28–30) was used to assemble a first version of the genome using both the ONT and WGS reads.

We followed the Arima-HiC mapping pipeline (https://github.com/ArimaGenomics/mapping_pipeline/blob/master/Arima_Mapping_UserGuide_A160156_v02.pdf) to prepare the data for

scaffolding. The pipeline aligns the sequencing data from each the Hi-C and Chicago dataset against the assembly from MaSuRCA, it then filters ligation adapters and removes PCR duplicates from the resulting alignments. These alignments were then processed with samtools (31,32) v1.13 and converted into BED files with bedtools (33) v2.30.

The MaSuRCA assembly was scaffolded with SALSA (34) v2.3 with ligation junction parameter -e AATT,GATC,CATG. Iteration number was set to 10 (-i 10) and we allowed for Hi-C/Chicago data to also correct assembly errors (-m yes).

We aligned the trimmed Illumina WGS reads to the scaffolded output of SALSA with bwa mem (35) v0.7.17-r1188 and used that alignment to polish the assembly with Pilon (36) v1.23. We repeated the alignment and polishing steps once. The error-corrected assembly was then screened for possible contaminants, using Blobtools2 (37) v3.1.0. Any contigs assigned to phyla other than Chordata were removed. However, any sequences categorized as 'No hits' were kept. The assembly was then manually curated by mapping the DpnII and MluCI Hi-C reads to the genome assembly with chromap (38) v0.2.2 with a quality filter of 0 and converted to a .hic file with Juicebox Assembly Tools (JBAT) (39) v2.14.00. Artisanal tools commit 9a79889 (https://bitbucket.org/bredeson/artisanal) was used to generate a JBAT

assembly file. We used the Juicebox GUI (40) v1.11.08 to manually curate the assembly with the .hic and .assembly files. Modifications made to the assembly included ordering and orienting scaffolds into chromosome-scale scaffolds, removing duplicated regions, and making manual assembly breaks to place misassembled contig pieces onto the correct scaffold. Artisanal was used to generate an updated genome assembly FASTA file. Scaffolds not placed on chromosomes were sorted by the strongest Hi-C connection to chromosome-scale scaffolds with genome assembly tools commit b0cda60 (https://github.com/conchoecia/genome_assembly_pipelines. D-Genies (41), accessed April 30th 2022, was used to align the manually-curated assembly to the chromosome-scale assembly of the closely related *Amphiprion percula* genome assembly (18). The evidence from this analysis was used to assign chromosome numbers to the *D. trimaculatus* scaffolds based on homology with *A. percula* chromosomes.

Genome quality assessment:

Benchmarking Universal Single-Copy Orthologs (BUSCO) (42,43) v5.2.2 was used to evaluate genome completeness by comparing number of orthologous genes found in the assembly to the 3,640 genes in the actinopterygii_odb10 database. Assembly statistics (assembly-stats;

https://github.com/sanger-pathogens/assembly-stats) were generated to track N50, L50, contigs, gaps, and lengths at each step. We used merqury (44) v1.3, to calculate the genome completeness and error rates.

Results

Sequencing

We sequenced four library types: a whole-genome shotgun library which resulted in 314.6Mb paired-end 150bp reads, representing 103x coverage, and 3.52M (4.84Gb) and 8.57M (19.77Gb) ONT reads from the two runs on the minION flowcells, representing 22x ONT coverage for the initial hybrid assembly. The five proximity ligation libraries used for scaffolding, two Hi-C (restriction enzymes DpnII and Mlucl), and three Chicago libraries (restriction enzymes DpnII, MlucI, and FatI) yielded ~108M, ~152M, ~65M, ~74M, ~67M, reads respectively for a total proximity ligation coverage of 154x. In total, across all data types, we had a final coverage of 280x (See Supplementary Table 1 for sequencing details).

Heterozygosity and repetitive sequence estimation

GenomeScope estimated the genome size to be 809 Mb, with 84% unique and 16% repetitive sequences, and 1.02% heterozygosity (Supplementary Figure 1).

Genome Assembly

Genome quality metrics for each step of the assembly are listed in Table 2. The initial *de novo* assembly by MaSuRCA with ONT and Illumina shotgun data had a total length of 919,275,268bp in 3,501 contigs with an N50 of 1,108Kb. Scaffolding with the HiC and Chicago libraries dropped the number of contigs to 2,467 and increased N50 to 16,013Kb. After two rounds of polishing with trimmed Illumina shotgun reads gaps decreased from 1,097 to 1,088. Blobtools2 showed that of the 2,467 contigs, none matched other taxa in NCBI databases of bacteria, invertebrates, mammals, phages, plants and fungi, or environmental samples. Four hundred seventy-eight contigs did not match any databases ('no-hits') and were left in the genome.

The manual curation of the genome assembly yielded 24 scaffolds consistent with chromosome-scale scaffolds (Figure 2). A dotplot comparison (Figure 3) with the *Amphiprion percula* (18) genome revealed that each of the *D. trimaculatus* chromosome-scale scaffolds had a one-to-one corresponding

homologous, albeit rearranged, chromosome in the *Amphiprion percula* genome.

The final assembly (GenBank accession: JAMOIN000000000) has a length of 910.7Mb, 90% of which was on chromosome-scale scaffolds and BUSCO score of 97.9%. Merqury calculated 86.19% completeness, QV of 44.6, and an estimated error rate 0.0000346, or a single nucleotide error every 28.9 Kb.

Discussion

The biology, evolution and biogeography of the three-spot damselfish is relatively well-studied using genetic (10,45–51) and genomic tools (52,53) and, as we shift further into the age of WGS data and tools, a reference genome is an invaluable resource. Here we present the chromosome-scale genome assembly of a three-spot damselfish, *Dascyllus trimaculatus,* collected from the Indonesian/Philippine population (20). It is the first within the genus *Dascyllus* of the widely studied, and large Pomacentridae family. This high-quality *de novo* assembly of a non-model coral reef fish is a valuable reference for furthering studies of evolutionary, ecological, and conservation studies for the species and for coral reef fish in general.

We report sequences for 24 chromosomes of the *D. trimaculatus* genome with total length and repetitive content (Figure 2A, Supplementary Figure 1) that is expected for this species (10,54,55). Interestingly, our Hi-C data also show that chromosomes three and four have strong connections at half the depth of other intra-chromosomal connections (Figure 2B). This pattern can be explained by a hemizygous state wherein one parental gamete contributed a Robertsonian-fusion of chromosomes three and four, and the other parental gamete contributed chromosomes three and four as separate chromosomes making the individual sequenced here, a 2n=47 individual. This finding is consistent with previous studies that report both 2n=47 and 2n=48 for *Dascyllus trimaculatus* (8,9,54). Chromosome numbers vary both within and among species of *Dascyllus.* One report on several *Dascyllus* species collected in the Philippines and the Ryukyu Archipelago of southern Japan demonstrated polymorphic karyotypes in all but one of the species (Ojima and Kashiwagi, 1981). *Dascyllus aruanus* had the most karyotypic variation - between 2n = 27-33 chromosomes, *D. reticulatus* 2n = 34-37, *D. trimaculatus* 2n = 47-48, and *D. melanurus* with 2n = 48.

In addition to confirming variation in chromosome number, the dot plot comparison between this genome and of the closest relative with an available chromosome-scale assembly, *Amphiprion percula* (18), revealed

several rearrangements in every chromosome between corresponding chromosomes (Figure 3). The Pomacentrid subfamilies Chrominae and Amphiprionini are estimated to have diverged over 50 million years ago (mya) (56). The estimated number of rearrangements within chromosomes ranged from 2+ in chromosome 7 of *D. trimaculatus* which was the most like its counterpart in *A. percula* to over 35 in chromosome 24 (Figure 3). This pattern of rearrangements has not been characterized between chromosome-scale genome assemblies of Pomacentridae. The role of variation in chromosome number has been the subject of several cytogenic studies which have found that chromosome diversity inversely related to mobility of the fish and that chromosome rearrangements can serve to either promote or prevent recombination events (11,12,14,57). Interestingly, chromosome 3 in the genome presented in this paper is one of the most rearranged while also being one of the chromosomes involved in the Robertsonian fusion mentioned above. This assembly will be a useful starting point to study how this type of genome structure varies at a meta-population scale, and how this influences recombination and adaptation.

This assembly represents the first chromosome-level genome of the genus *Dascyllus* as well as the first non-*Amphiprion* chromosome-scale genome published in the Pomacentridae family. Damselfishes are excellent

model species due to their relatively small size, ease to manage in the wild and lab, and those interested in this group will benefit from this addition to the available genomic resources. *Dascyllus trimaculatus* itself, is has had a dynamic evolutionary trajectory across the Indo-Pacific, evident in species complex that is continuing to reveal its complexity and provide insight into evolutionary mechanisms. In addition to providing a high-quality reference genome to further our understanding of genomic architecture, this assembly will serve to leverage information stored across the genome to better understand the population dynamics, phylogeny, biogeography, demographics, of *Dascyllus trimaculatus*, as well as gain insight into historical, current, and future response to changes in climate.

Data Availability

The assembly and genomic sequencing reads generated for this study have all been deposited in the NCBI GenBank database under BioProject ID PRJNA828170. Accession numbers for the genome and data are listed in Supplementary Table 1 on page 70.

Acknowledgements

The land on which much of this work was conducted is the unceded territory of the Awaswas-speaking Uypi Tribe. The Amah Mutsun Tribal Band, comprised of the descendants of indigenous people taken to missions Santa Cruz and San Juan Bautista during Spanish colonization of the Central Coast, is today working hard to restore traditional stewardship practices on these lands and heal from historical trauma.

The authors are also grateful for access to the Hummingbird computational cluster and the team behind it at University of California Santa Cruz**.** The whole genome shotgun sequencing was carried out at the DNA Technologies

and Expression Analysis Core at the University of California Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. We also appreciate Jonas Oppenheimer and Robert Lehmann for helpful discussions and advice.

Conflict of Interest

None declared.

Funding

This work was supported in part by Marilyn C. Davis Scholarship and American Association of University Women (through STARS Reentry program at UCSC), Burnand-Partridge Foundation, tuition support in part by a Dissertation Year (1 quarter) Fellowship (UCSC) to M.B.R.

References

1. Hiatt RW, Strasburg DW. Ecological Relationships of the Fish Fauna on Coral Reefs of the Marshall Islands. Ecol Monogr. 1960;30(1):65–127.

2. Allen GR. Damselfishes of the world [Internet]. Melle, Germany; Mentor, Ohio: Mergus ; Aquarium Systems [distributor]; 1991. Available from: http://catalog.hathitrust.org/api/volumes/oclc/24436027.html

3. Allen GR, Werner TB. Coral Reef Fish Assessment in the 'Coral Triangle' of Southeastern Asia. Environ Biol Fishes. 2002 Oct;65(2):209–14.

4. Bellwood DR, Wainwright PC. Chapter 1: The History and Biogeography of Fishes on Coral Reels. In: Coral Reef Fishes. Elsevier Science; 2002. p. 28.

5. Tang KL, Stiassny MLJ, Mayden RL, DeSalle R. Systematics of Damselfishes. Ichthyol Herpetol [Internet]. 2021 May 5 [cited 2022 May 17];109(1). Available from: https://bioone.org/journals/ichthyology-and-herpetology/volume-109/issue-1/i2020105/Systematics-of-Damselfishes/10.1643/i2020105.full

6. Wellington GM, Victor BC. Planktonic larval duration of one hundred species of Pacific and Atlantic damselfishes (Pomacentridae). Mar Biol. 1989;101(4):557–67.

7. Robitzch VSN, Lozano-Cortés D, Kandler NM, Salas E, Berumen ML. Productivity and sea surface temperature are correlated with the pelagic larval duration of damselfishes in the Red Sea. Mar Pollut Bull. 2016 Apr;105(2):566–74.

8. Ojima Y, Kashiwagi E. Chromosomal evolution associated with Robertsonian fusion in the genus Dascyllus (Chrominae, Pisces).: Chrominae, Pisces. Proc Jpn Acad Ser B. 1981;57(10):368–70.

9. Kashiwagi E, Takai A, Ojima Y. Chromosomal Distribution of Constitutive Heterochromatin and Nucleolus Organizer Regions in Four Dascyllus Fishes (Pomacentridae, Perciformes). Cytologia (Tokyo). 2005;70(3):345–9.

10. Getlekha N, Supiwong W, Yeesin P, Pengseng P, Kasiroek W, Tanomtong A. Chromosomal Characteristics of the Three-Spot Damselfish, *Dascyllus trimaculatus* (Perciformes, Pomacentridae) in Thailand. Cytologia (Tokyo). 2017;82(1):51–7.

11. Molina WF, Galetti PM. Karyotypic changes associated to the dispersive potential on Pomacentridae (Pisces, Perciformes). J Exp Mar Biol Ecol. 2004 Sep;309(1):109–19.

12. Galetti PM, Aguilar CT, Molina WF. An overview of marine fish cytogenetics. In: Solé-Cava AM, Russo CAM, Thorpe JP, editors. Marine Genetics [Internet]. Dordrecht: Springer Netherlands; 2000 [cited 2022 May 16]. p. 55–62. Available from: http://link.springer.com/10.1007/978-94-017-2184-4_6

13. Hardie DC, Hebert PD. Genome-size evolution in fishes. Can J Fish Aquat Sci. 2004 Sep 1;61(9):1636–46.

14. Martinez PA, Zurano JP, Amado TF, Penone C, Betancur-R R, Bidau CJ, et al. Chromosomal diversity in tropical reef fishes is related to body size and depth range. Mol Phylogenet Evol. 2015 Dec;93:1–4.

15. Hotaling S, Kelley JL, Frandsen PB. Toward a genome sequence for every animal: Where are we now? Proc Natl Acad Sci. 2021 Dec 28;118(52):e2109019118.

16. Tan MH, Austin CM, Hammer MP, Lee YP, Croft LJ, Gan HM. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (Amphiprion ocellaris) genome assembly. GigaScience. 2018 Jan 12;7(3):gix137.

17. Schunter C, Welch MJ, Ryu T, Zhang H, Berumen ML, Nilsson GE, et al. Molecular signatures of transgenerational response to ocean acidification in a species of reef fish. Nat Clim Change. 2016 Nov;6(11):1014.

18. Lehmann R, Lightfoot DJ, Schunter C, Michell CT, Ohyanagi H, Mineta K, et al. Finding Nemo's Genes: A chromosome-scale reference assembly of the genome of the orange clownfish Amphiprion percula. Mol Ecol Resour. 2019;19(3):570–85.

19. Ryu T, Herrera M, Moore B, Izumiyama M, Kawai E, Laudet V, et al. A chromosome-scale genome assembly of the false clownfish, Amphiprion ocellaris. G3 Genes Genomes Genet. 2022 Mar;00(0):12.

20. Limon J, Roberts MB, Schultz DT, Bernardi G. The complete mitochondrial genome of Dascyllus trimaculatus. Mitochondrial B. *in review*

21. Dawson MN, Raskoff KL, Jacobs DK. Field preservation of marine invertebrate tissue for DNA analyses. Mol Mar Biol Biotechnol. 1998;7(2):145–52.

22. Sambrook, Joseph, and David W. Russell. Purification of Nucleic Acids by Extraction with Phenol:chloroform. CSH Protoc. 2006;

23. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 2016 Mar;26(3):342–50.

24. Adams M, McBroome J, Maurer N, Pepper-Tunick E, Saremi NF, Green RE, et al. One fly–one genome: chromosome-scale genome assembly of a single outbred Drosophila melanogaster. Nucleic Acids Res. 2020 Jun 3;gkaa450.

25. Guppy Basecalling Software [Internet]. Oxford Nanopore Technologies; 2019. Available from: https://nanoporetech.com/

26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

27. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Berger B, editor. Bioinformatics. 2017 Jul 15;33(14):2202–4.

28. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013 Nov;29(21):2669–77.

29. Jiang JB, Quattrini AM, Francis WR, Ryan JF, Rodríguez E, McFadden CS. A hybrid *de novo* assembly of the sea pansy (*Renilla muelleri*) genome. GigaScience [Internet]. 2019 Apr 1 [cited 2022 Apr 21];8(4). Available from:

https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giz026/5426564

30. Wang W, Das A, Kainer D, Schalamun M, Morales-Suarez A, Schwessinger B, et al. The draft nuclear genome assembly of Eucalyptus pauciflora: a pipeline for comparing de novo assemblies. GigaScience. 2020 Jan 1;9(1):giz160.

31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

32. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021 Jan 29;10(2):giab008.

33. Quinlan AR, Clark RA, Sokolova S, Leibowitz ML, Zhang Y, Hurles ME, et al. Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Res. 2010 May 1;20(5):623–35.

34. Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. Ioshikhes I, editor. PLOS Comput Biol. 2019 Aug 21;15(8):e1007273.

35. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

36. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. Wang J, editor. PLoS ONE. 2014 Nov 19;9(11):e112963.

37. Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies. F1000Research. 2017 Jul 31;6:1287.

38. Zhang H, Song L, Wang X, Cheng H, Wang C, Meyer CA, et al. Fast alignment and preprocessing of chromatin profiles with Chromap. Nat Commun. 2021 Dec;12(1):6566.

39. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom | Elsevier Enhanced Reader. Cell Syst. 2016 Jul 27;3:99–101.

40. Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, et al. The Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under $1000 [Internet]. Genomics; 2018 Jan [cited 2022 May 16]. Available from: http://biorxiv.org/lookup/doi/10.1101/254797

41. Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 2018 Jun 4;6:e4958.

42. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015 Oct 1;31(19):3210–2.

43. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol Biol Evol. 2018 Mar 1;35(3):543–8.

44. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020 Dec;21(1):245.

45. Bernardi G, Crane NL. Molecular phylogeny of the humbug damselfishes inferred from mtDNA sequences. J Fish Biol. 1999;54(6):1210–7.

46. Bernardi G, Holbrook SJ, Schmitt RJ. Gene Flow at three spatial scales in a coral reef Fish, the three-spot dascyllus, Dascyllus trimaculatus. Mar Biol. 2001;138:457–65.

47. McCafferty S, Bermingham E, Quenouille B, Planes S, Hoelzer G, Asoh K. Historical biogeography and molecular systematics of the Indo-Pacific genus Dascyllus (Teleostei: Pomacentridae). Mol Ecol. 2002;11(8):1377–92.

48. Leray M, Beldade R, Holbrook SJ, Schmitt RJ, Planes S, Bernardi G. Isolation and characterization of 13 polymorphic nuclear microsatellite primers for the widespread Indo-Pacific three-spot damselfish, *Dascyllus*

trimaculatus , and closely related *D. auripinnis*. Mol Ecol Resour. 2009 Jan;9(1):213–5.

49. Leray M, Beldade R, Holbrook SJ, Schmitt RJ, Planes S, Bernardi G. Allopatric Divergence and Speciation in Coral Reef Fish: The Three-Spot Dascyllus, *Dascyllus trimaculatus*, Species Complex. Evolution. 2010;64(5):1218–30.

50. Liggins L, Treml EA, Possingham HP, Riginos C. Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. J Biogeogr. 2016;43(2):256–67.

51. Crandall ED, Riginos C, Bird CE, Liggins L, Treml E, Beger M, et al. The molecular biogeography of the Indo-Pacific: Testing hypotheses with multispecies genetic patterns. Glob Ecol Biogeogr. 2019;28(7):943–60.

52. Salas EM, Bernardi G, Berumen ML, Gaither MR, Rocha LA. RADseq analyses reveal concordant Indian Ocean biogeographic and phylogeographic boundaries in the reef fish *Dascyllus trimaculatus*. R Soc Open Sci. 2019 May;6(5):172413.

53. Salas EM, Hobbs JA, Bernal MA, Simison WB, Berumen ML, Bernardi G, et al. Distinct patterns of hybridization across a suture zone in a coral reef fish (*Dascyllus trimaculatus*). Ecol Evol. 2020 Mar;10(6):2813–37.

54. Arai R. Chromosomes of four species of coral fishes form Japan. Bull Natn Sci Mus Tokyo Ser A. 1976;2:137–41.

55. Yuan Z, Liu S, Zhou T, Tian C, Bao L, Dunham R, et al. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. BMC Genomics. 2018 Dec;19(1):141.

56. McCord CL, Nash CM, Cooper WJ, Westneat MW. Phylogeny of the damselfishes (Pomacentridae) and patterns of asymmetrical diversification in body size and feeding ecology. PLOS ONE. 2021 Oct 27;16(10):e0258889.

57. Kirkpatrick M, Barton N. Chromosome Inversions, Local Adaptation and Speciation. Genetics. 2006 May 1;173(1):419–34.

Ch2 - Figures and Tables

Figure 1. *Dascyllus trimaculatus* and species complex global distribution. The three-spot damselfish (Pacific morph) is shown on the right of the map adapted from Leray et al. (2009) and Salas et al. (2020). The map shows in blues the broad distribution of *D. trimaculatus*. Differences in blues reflect results of Salas et al. (2020) which showed Indian Ocean differentiation from Pacific populations as well as a sub population in Cocos Keeling and a hybridization zone in Christmas Island. Similarly, the darker blue patch in the Central Pacific shows another divergent population of *D. trimaculatus* identified in Leray et al (2009). The other colors show the distributions of the other species within the *D. trimaculatus* complex: green for *D. albisella*, yellow for *D. auripinnis*, and red for *D. strasburgi.*

Figure 1.

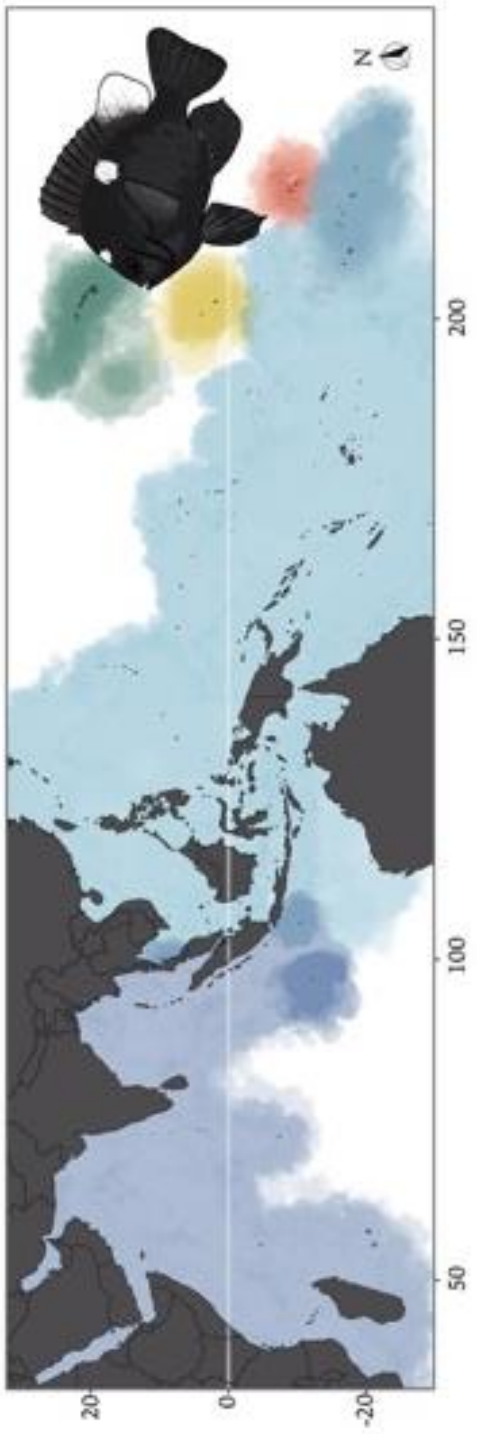


Figure 2. Genome statistics and chromosome map.

Panel A: The outer circumference of the main plot represents the full length of the 910,763,285 bp chromosome-scale assembly of *Dascyllus trimaculatus*. The outer ring of blues depicts GC (dark blue) and AT (light blue) content along the assembly which is summarized in the lower left. The second ring is demarcated by percentage of the total contigs of the genome. Orange and pale-orange arcs show the N50 and N90 record lengths (34,909,338 and 22,085,708 bp), respectively overlying the dark grey, which arranges scaffolds in order by size starting from the largest scaffold (41,400,476 bp and ~4% genome, shown in red). The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. A summary of BUSCO statistics for complete (97.9%), fragmented (0.7%), duplicated (1.7%), and missing (1.5%), orthologous genes in the actinopterygii_odb10 set is shown in the top right. **Panel B:** A Hi-C contact map made with the MluCI and the DpnII libraries showing 24 chromosome clusters and the unscaffolded contigs. In the green square, chromosomes 3 and 4 show strong interchromosomal connections at roughly half coverage indicating Robertsonian fusion in one set of chromosomes contributed a parent with 2n=47 while the other parent contributed 2n=48.

Figure 2.

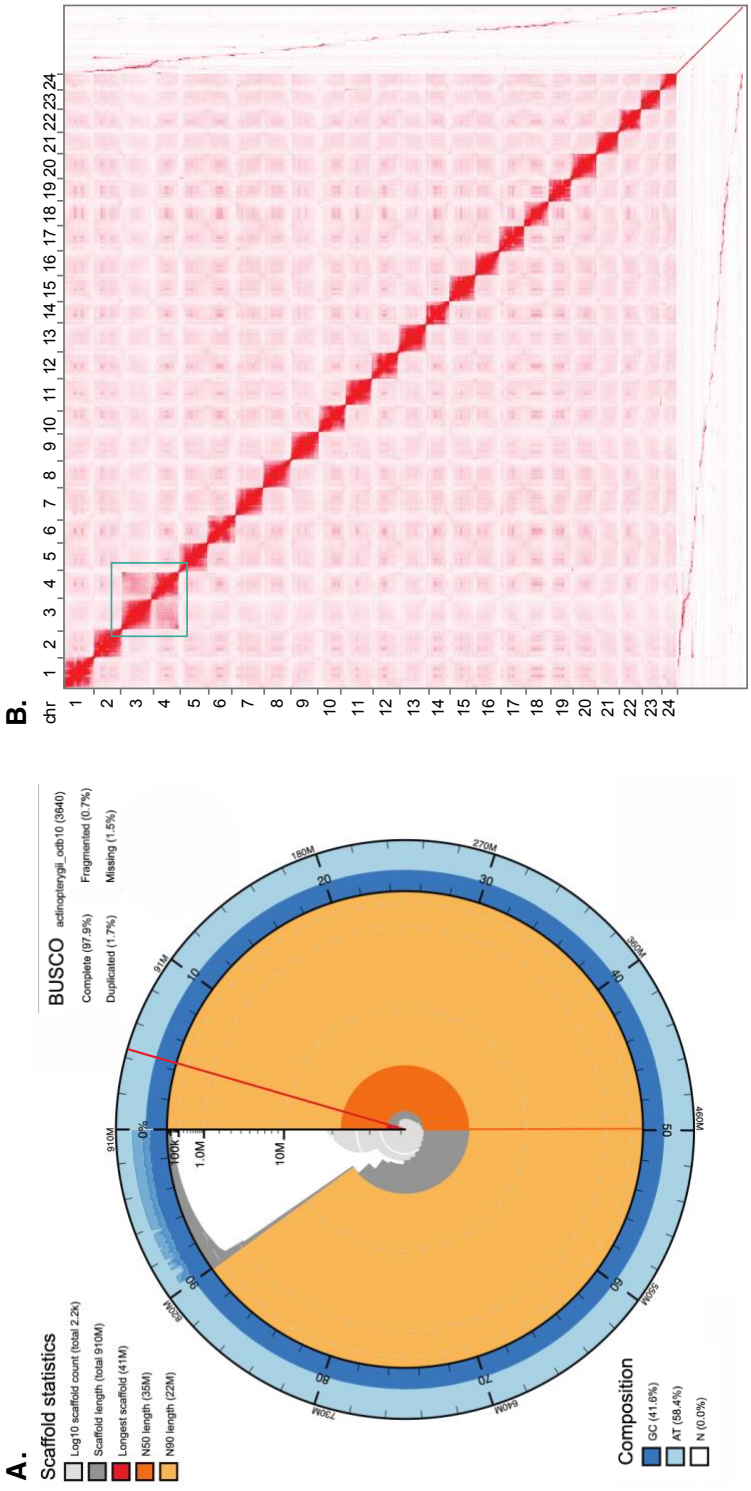

B.
chr
A.
Scaffold statistics
Log10 scaffold count (total 2.2k)
Scaffold length (total 910M)
Longest scaffold (41M)
N50 length (35M)
N90 length (22M)
BUSCO aclinoptenygii_odb10 (3840)
Complete (97.9%)
Fragmented (0.7%)
Duplicated (1.7%)
Missing (1.5%)
Composition
GC (41.8%)
AT (58.4%)
N (0.0%)

Figure 3. ***Dascyllus trimaculatus*** **mapped against chromosome-level genome of** ***Amphiprion percula*** **(Pomacentridae)**

The main panel shows a dot-plot of the assemblies of *Dascyllus trimaculatus* (presented in this manuscript) and the genome of the anemone fish *Amphiprion percula* (Lehmann, 2019). The remaining un-scaffolded contigs are shown in the last column. The right three panels show a close-up dot plot of the color-coded boxes in the main panel of chromosome 3, chromosome 7, and chromosome 24. Chromosome 3, one of the chromosomes involved in Robertsonian fusions within the species, shows many rearrangements as well as regions of repeat sequences near one end. Chromosome 7 seems to be one of the most architecturally conserved chromosomes between *Dascyllu*s and *Amphiprion,* whereas Chromosome 24 shows and examples of a highly rearranged chromosome.

Figure 3.

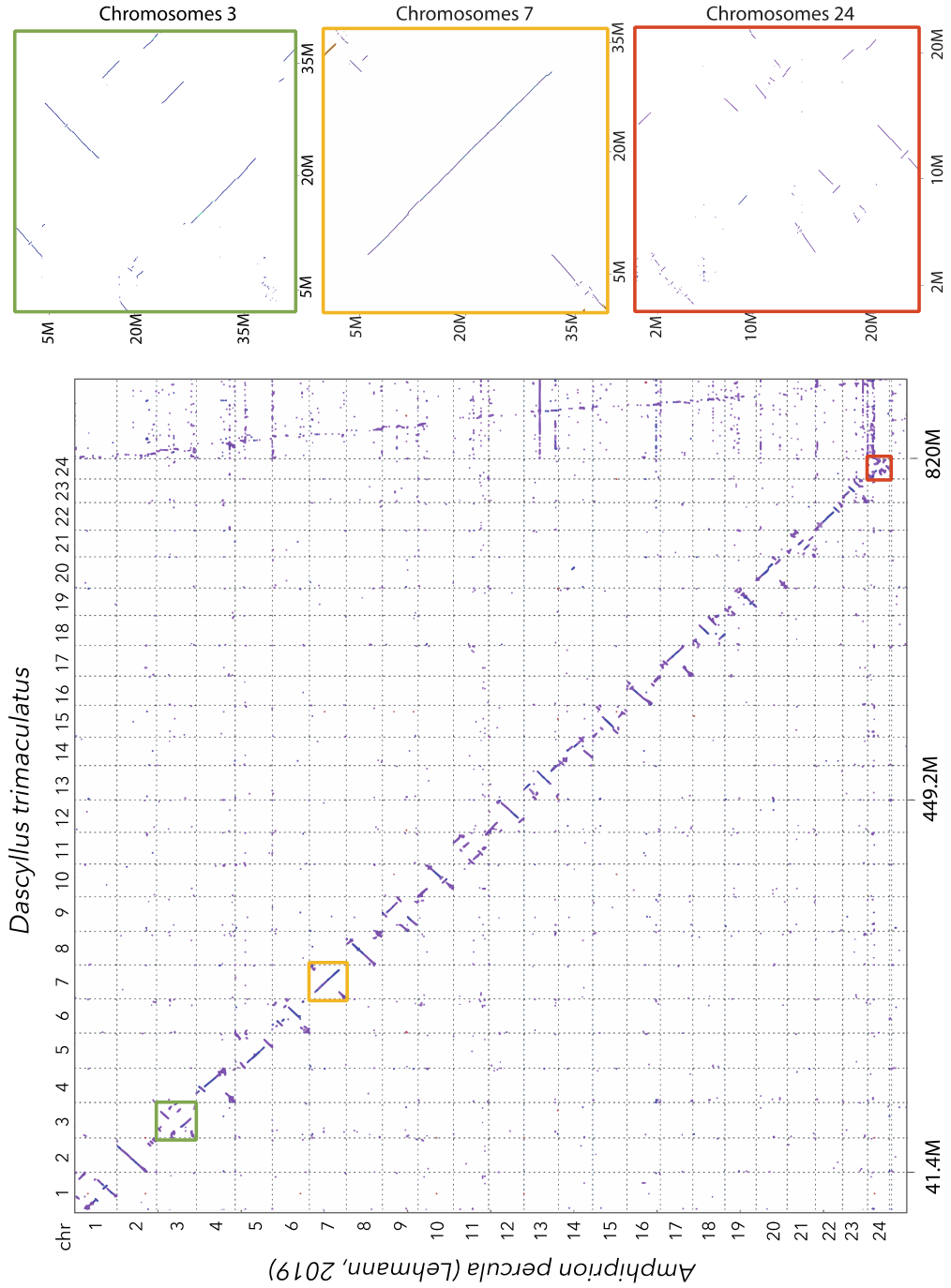
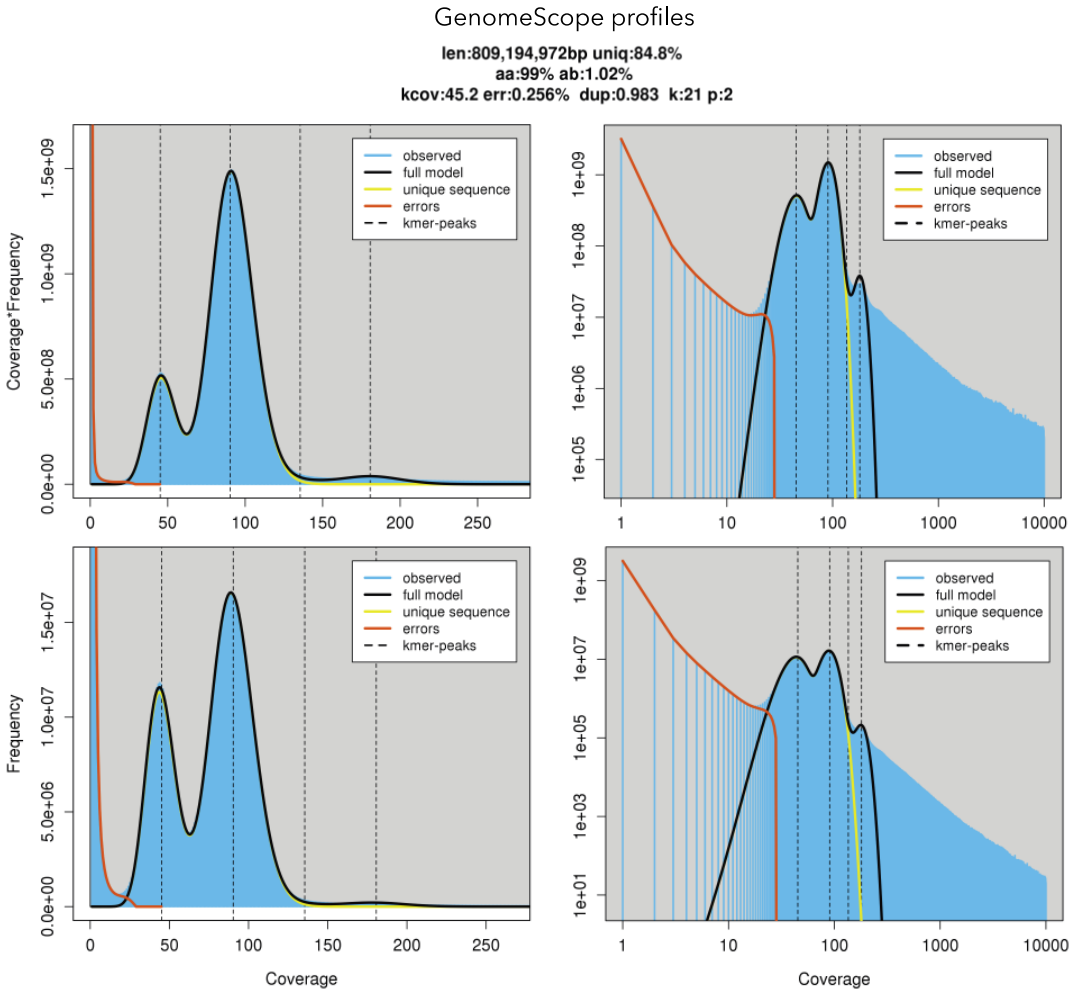

Chromosomes 3
Chromosomes 7
Chromosomes 24
5M
20M
35M
5M
20M
35M
2M
10M
20M
Dascyllus trimaculatus
Amphiprion percula (Lehmann, 2019)
chr
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
41.4M
449.2M
820M

Figure S1. GenomeScope results from paired-end 150bp whole genome sequences of *Dascyllus trimaculatus.*

Supplementary Figure 1.



property	min	max
Homozygous (aa)	98.9744%	98.983%
Heterozygous (ab)	1.01705%	1.02557%
Genome Haploid Length	808,752,873 bp	809,194,972 bp
Genome Repeat Length	122,968,760 bp	123,035,980 bp
Genome Unique Length	685,784,113 bp	686,158,993 bp
Model Fit	86.9621%	97.8826%
Read Error Rate	0.256184%	0.256184%

Table 1.
List of programs and program versions in order of use for the genome assembly of the three-spot damselfish, *Dascyllus trimaculatus*

Table 1.

Purpose		Software and version
Estimate genome size		Jellyfish v2.2.10
Reference-free characterization		GenomeScope v2
Basecalling		Guppy v3.3
De novo assembly		MaSuRCA downloaded Sept 2020
Map HiC/Chicago reads		Arima Hi-C pipeline Release date: 05. 2019 BWA v0.7.17-r1188; samtools v1.13; Picard v2.26.2
File processing		Samtools v1.13, bedtools v2.30.0
Proximity ligation scaffolding		SALSA2 v2.3
Trim adapters and filter		Trimmomatic v0.39
Polish assembly (2x)		Pilon v1.23
Check for contamination		Blobtools2 v3.1.0
Manual curation	Map Hi-C to genome	Chromap v0.2.2
	File conversion	Juicebox Assembly Tools v2.14.0
		Artisinal tools 9a79889
		https://github.com/conchoecia/genome_assembly_pipelines b0cda60
	Chromosome assignment	D-Genies Accessed April, 2022
Assembly statistics	Completeness	BUSCO v5.2.2, assembly-stats v0.0.1, merqury v1.3
	General stats	Assembly-stats v0.0.1
	Quality and error rate	Merqury v1.3

Table 2.
A comparison of genome metrics between *Dascyllus trimaculatus* assembly stages.

Table 2.

		Assembly Version				
		MaSuRCA	SALSA2	Pilon 1st iteration	Pilon 2nd iteration	Final assembly
Assembly step		De novo hybrid assembly	Scaffolding-proximity ligation	Error correcting	Error correcting	Manual assembly curation
Input data		ONT + Illumina shotgun	Hi-C + Chicago	Illumina shotgun	Illumina shotgun	Hi-C dpnII and mluCI data
Genome length (bp)		919 275 268	919 814 768	919 296 307	919 364 398	910 763 285
Number of contigs		3 501	2 467	2 467	2 467	2 156
Average contig length (bp)		262 575.06	372 847.49	372 637.34	372 664.94	422 431.95
Largest contig length (bp)		8 494 454	34 904 224	34 891 597	34 891 597	41 400 476
N50 length : n		1 108 072 : 200	16 013 467 : 20	16 006 006 : 20	16 006 752 : 20	34 909 338 : 13
N90 length : n		124 315 : 1 137	218 694 : 274	218 662 : 274	218 658 : 274	22 085 708 : 24
N100 length : n		1 790 : 3 501	1 790 : 2 467	1 790 : 2 467	1 790 : 2 467	1790 : 2 156
Gaps		18	1 097	1 090	1 088	1 411
N count		1 800	541 300	471 149	448 527	141 100
BUSCO scores	C	97.3	97.9	97.9	97.9	97.9
(% of 3 640 orthologs)	S	94.5	95.7	95.8	95.7	96.2
	D	2.8	2.2	2.1	2.2	1.7
	F	0.5	0.6	0.6	0.6	0.7
	M	2.2	1.5	1.5	1.5	1.4
Completeness (%)		86.18	86.18	86.16	86.19	85.96
Quality value		44.7342	44.7342	44.2101	44.6053	44.61
Error rate		0.000033619	0.000033619	3.79307E-05	0.000046314	0.000034597

BUSCO scores. (C)omplete and (S)ingle; (C)omplete and (D)uplicated; (F)ragmented and (M)issing BUSCO genes. n, number of BUSCO genes in the set/database.
Bp: base pairs.

Table S1.
Sequencing information of data used to assemble the genome of the three-spot damselfish, *Dascyllus trimaculatus*, Bioproject: PRJNA828170 Biosample: SAMN27642109; Genome accession: JAMOIN000000000 for isolate Kuro_0920G

Table S1.

Purpose	Data Type	Linker Sequence	Sequencing Chemistry	Sequencing Institution	Number of bases	Coverage	GenBank accession
High accuracy short reads	NEB Next Ultra II WGS		HiSeq 4000 150PE	UC Davis	94383867000	103.6x	SRX17663068
Proximity ligation	Chicago - DpnII	GATCGATC	HiSeq 4000 150PE	Fulgent Genetics	19712640600	21.6x	SRX17663071
	Chicago - MluCI	AATTAATT			22305540600	24.5x	SRX17663072
	Chicago - FatI	CATGCATG			20104123200	22.1x	SRX17663073
	HiC - DpnII	GATCGATC			32486064300	35.7x	SRX17663069
	HiC - MluCI	AATTAATT			45729378900	50.2x	SRX17663070
Long-reads for scaffolding	ONT-MinION		SQK-LSK109/FlowCell 9.4.1	UCSC	4789005362	5.3x	SRX17742644
					15220684630	16.7x	SRX17742645

Chapter 3

Population genomics of edge populations in the three-spot damselfish, *Dascyllus trimaculatus* based on whole genome sequencing

May Roberts[1], Angela Quieros[2], Jayvee Sacco[3], Liam Zarri[4], Giacomo Bernardi[1]

1. Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, California, USA
2. Akkeshi Marine Station, Hokkaido University, Hokkaido, Japan
3. Verde Island Passage Center for Oceanographic Research and Aquatic Life Sciences (VIP CORALS), Batangas State University, Lobo Campus, Masaguitsit, Lobo, Batangas Philippines
4.Department of Natural Resources, Cornell University, Ithaca, New York, USA

Abstract

The three-spot damselfish, (*Dascyllus trimaculatus)* is a small, common, and widespread coral reef fish that is well studied both at the ecological and at the molecular level in large part due to its membership in a species complex that spans the Indo Pacific. In this study, low coverage whole genome sequencing is used to examine the population structure of its Pacific distribution, and at a more fine-scale, along the Ryukyu Archipelago in southern Japan. We find evidence of clear divergence between the Pacific and Indian Ocean populations as well as between the Pacific and the restricted edge population in the Society Islands at the eastern edge of its range. We see a clear signal of introgression of *D. auripinnis* described from the Line and Phoenix islands into *D. trimaculatus* that extends across to the west Pacific - further than previously reported. We also find that there is directional but restricted gene flow of Philippine genotypes into the Ryukyu Archipelago. While we have not yet identified specific isolation and speciation mechanisms, we propose some likely scenarios. Finally, this study adds further information, genomic resources, and questions to pursue for this model species and system.

Introduction

Indo-Pacific coral reef fishes offer a system to explore fundamental concepts of both theoretical and applied population genomics. Indeed, coral reef fishes have been used to study evolution, phylogenetics, phylogeography, speciation, behavior, and population dynamics. They are also the primary source of proteins for millions of people, primarily in developing countries, which underscores the crucial role that population genomics may play in conservation and management of essential resources. In the past, a number of model species have been used to tackle such questions, and the presence of widely distributed species has been shown to provide a perfect setup.

The three-spot damselfish, *Dascyllus trimaculatus,* has been studied extensively both at the ecological and at the molecular level. This relatively small damselfish (family Pomacentridae) recruits and lives in anemones as a juvenile, but unlike anemonefishes, leaves the protection of the anemones when it reaches a size large enough to avoid predation. As an adult, it lives a demersal life, usually in the vicinity of the anemones, feeding on plankton. Adults reproduce several times a year by spawning on rocks where males will guard eggs for several days. After hatching, larvae have a pelagic phase before they recruit back to a reef into an anemone (pelagic larval duration, PLD, lasts approximately 23-30 days)(Wellington and Victor 1989). These life history characteristics translate as populations that are genetically structured, as has been shown in previous studies (Giacomo Bernardi, Holbrook, and Schmitt 2001; Leray et al. 2010; Salas et al. 2020; McCafferty et al. 2002). In fact, *D.*

trimaculatus belongs to a species complex that is geographically widespread. *Dascyllus trimaculatus* is found across the Indo Pacific, from the Red Sea to French Polynesia in the South and Japan in the North of the Pacific Ocean. Yet, within this vast distribution, closely related species, with much smaller distributions, have been described. Two species are restricted to remote island groups – the Marquesas (*D. strasburgi*) and the Hawaiian Archipelago and Johnston Atoll (*D. albisella,)*. *Dascyllus auripinnis*, is another species within the complex that has been described from the Line Islands and adjacent atolls. In addition to these described species, population genetic studies based on mitochondrial markers and nuclear microsatellites have shown strong population structure in *D. trimaculatus,* where a very strong genetic barrier is found between Indian Ocean and Pacific populations, and again within the Pacific Ocean, separating French Polynesian individuals as its own group. A recent study (Chapter 1) based on more than 600 individuals found that the Society Islands of French Polynesia and the Ryukyu Archipelago, two marginal populations, generally separated out from the greater Pacific cluster into their own clades (though some from each location did fall into other clades (Chapter 1). These findings raised the question of the origin, nature, and role of edge populations as these two are found at the extreme southern and northern edges of the distribution of the species, respectively. Each offer a unique example of what edge populations look like and the conditions under which they persist.

The goal of this study was to use a more powerful approach, low coverage whole genome sequencing, to first confirm the presence of multiple species in the *D.*

trimaculatus species complex, second: to identify gene flow boundaries within the Pacific Ocean, and finally, to determine the status of edge populations and their genetic relationships with the center of distribution of the species.

Methods

Tissue collection and Sampling locations

This study includes 9-14 individuals for each of 12 locations across the species complex range including three locations in the Ryukyu Archipelago. This includes the Ryukyu Archipelago (Amami-Oshima, Okinawa, Iriomote), Philippines (Verde Island Passage), Micronesia (Falalop, Ulithi), Fiji (Tavenui), Baker Atoll, Kingman Reef, French Polynesia (Mo'orea), Marquesas, Hawai'ian Islands (Hawai'i) and the Indian Ocean outgroup Oman (Umm al Ghanam, Musandam) (see Table 1 for more details). Fin and/or gill tissue for the 129 individuals *D. trimaculatus* was collected using pole spears or hand nets while scuba or free diving and preserved in 95% ethanol. Sampling protocols were approved by Institutional Animal Care and Use Committee (IACUC) at University of California Santa Cruz.

Whole-genome library preparation and sequencing:

Total genomic DNA was extracted by first lysing all tissue by digesting overnight in lysis buffer, 20% SDS, and proteinase K, before purifying using a chloroform and ethanol precipitation protocol, treated with RNase and eluted in purified water (See Supplemental Protocol 1). DNA was then size selected using beads to remove

fragments less than 1000bp. Whole genome libraries were prepared following a Tn5 transposase protocol where DNA was tagmented and indexed (some made by the Corbett-Detig Lab UCSC and the rest by Nextera UDI) byt the Tn5 transposase and PCR amplified (additional detail in Supplemental Protocol 2). The final libraries were bead cleaned aimed to remove fragments greater than 700bp. These were then checked by measuring DNA concentration using Qubit and a representative few were run on a Bioanalyzer to check fragment distribution. Libraries were based on concentration and 150bp were paired-end sequenced on 4 partial lanes of a NovaSeq (NovaGene, Sacramento, CA).

Data-processing:

Details and scripts used for data processing and all analysis can be found at: https://github.com/mayroberts/lcwgs-population-analysis/. Paired-end fastq.gz files were adapter and polyG tail trimmed, put through initial filtering based on default parameters for low quality reads and read length (<30bp removed) using *fastp* v0.12.4 (Chen et al. 2018) and resulting pre and post qc reports visualized with *multiqc* v1.11 (Ewels et al. 2016). Reads were mapped to the *D. trimaculatus* genome (Roberts et al in review) that had been filtered of contigs <500 bp and resulting bam files were filtered to remove non-uniquely mapped reads and those with a low quality score (q 20) with *bowtie2* v2.4.5 (Langmead and Salzberg 2012) and *samtools* v1.13(Danecek et al. 2021; Li et al. 2009). Final coverage per individual and total was calculated using a combination of *samtools* and a custom R script adapted

from (https://github.com/therkildsen-lab/data-processing/blob/master/lcwgs_data_processing.md). This was used to determine that minimum and maximum global single nucleotide polymorphism (SNP) depth filters would be set at 100 and 590 respectively.

Global SNP calling was done with *angsd* v0.939(Korneliussen, Albrechtsen, and Nielsen 2014) with –minind 13 (set to 10% of total individuals), --minq 20 (minimum base quality score), --minmaf 0.05 (minimum minor allele frequency), --minmapq 20 (minimum mapping quality score), --mindp 100 and –maxdp 590 as stated before. The samtools genotype likelihood model (--gl 1) within *angsd* was chosen to calculate minor allele frequencies. Minor allele frequency estimations for populations were also calculated similarly except –minind 6(2/3 the minimum population size), --mindp 6 (same as –minind), and –maxdp 14 (minind x estimated average coverage (2.3)).

Analysis:

To calculate pairwise weighted Fst, first *angsd* was used to estimate site allele frequency likelihood for each population or grouping (--doSaf 1). These pairwise comparison were between each sampling location as well as by major clades defined by the mtDNA phylogenetic tree in Chapter 1 (*D. albisella, D. strasburgi*, *D. trimaculatus* - Indian Ocean, *D. trimaculatus* Mo'orea, *D. trimaculatus* Pacific Rim and *D. auripinnis*). The 2D site frequency spectrum was estimated (angsd realSFS) and the weighted Fst average was calculated (realSFS fst stat). For details on

parameters and scripts used for how pair-wise Fsts were obtained for populations and by taxonomic and other groupings using *angsd*, see https://github.com/mayroberts/lcwgs-population-analysis-minus-the-cursing/blob/main/Fst.md.

Global genotype likelihoods were used to estimate individual admixture proportions in the program NGSadmix for k 4-10 using the default parameters and -seed 4 (Skotte, Korneliussen, and Albrechtsen 2013). R was then used to plot the resulting proportions.

Results

The 129 individuals in this study received on average 2.3x coverage. The mean depth per position across all individuals is 282.68 (sd=160.5406). A total of 867,185,962 sites (95.22 % of the reference genome) were covered at least once. After filtering (--mindp 100 and –maxdp 590), a total of 866,405,756 sites were analyzed and 22,573,846 snps were retained and used in further analysis.

All average weighted Fsts for population-wise comparisons are reported in Table 2a and Fsts calculated between major clades are reported in Table 2b. Fst generally increased with distance between groups. Lowest fsts were between the Ryukyu locations which ranged between 0.027-0.029. Gene flow was more limited among populations within regions in the West Pacific where Fsts ranged between 0.027 (Okinawa/Amami-Oshima) and 0.058 (Ulithi/Philippines). Populations assigned to the recently described *D. auripinnis,* collected at Baker Island and

Kingman Atoll showed restricted gene flow with populations assigned to other species within the complex, but showed some gene flow with some adjacent populations such as Fiji and Ulithi, Fst=0.057 -0.073. Fst values were greater than 0.1 for pairwise comparisons with the other western Pacific populations. Samples from Mo'orea (*D. trimaculatus*), Marquesas (*D. strasburgi*), Hawai'i (*D. albisella*), and Oman (*D. trimaculatus*. – Indian Ocean) all showed low gene flow (Fst=0.17 - 0.42) between themselves and all other locations.

Table 2b shows a similar pattern where the major clades have little gene flow between them except for between the Pacific Rim and Baker Atoll and Kingman Reef (Fst= 0.005). Admixture analysis (Figure 2) showed that most locations (Oman, Marquesas, Hawai'i, Mo'orea, Baker Atoll, Kingman Reef, Philippines) had very little, to no signature of admixture with other groups. However, the Ryukyu populations, Ulithian, and Fijian populations showed varying levels of admixture, which showed some proportion of *D. auripinnis* at decreasing levels with distance from Baker Atoll and Kingman Reef. Figure 3 is a subset of Figure 2 showing a genetic cline along a gradient of four populations along the Kuroshio Current: Philippines, Iriomote, Amami-Oshima.

Discussion

While much genetic information has been gathered to understand the geographic distribution of genotypes in the *Dascyllus trimaculatus* species complex, this study, which harnesses the power of low coverage whole genome sequencing sheds new

light on the genetic architecture of this species complex. Here, we present the first Pacific-wide analysis of population structure of the *D. trimaculatus* complex.

Species designations

Current species designations within the *D. trimaculatus* complex are due for an update. From previous work using mitochondrial markers, microsatellites, RADseq and this study based on low coverage whole genome sequencing, the species *D. trimaculatus*, is likely composed of at least three separate species; the one originally described from the Indian Ocean (Rüppell, 1829), the main Pacific clade, and the more restricted population found in the Society Islands. Previous genetic studies which include samples from these regions show these clear divergences (Leray et al. 2010; Salas et al. 2020; Giacomo Bernardi, Holbrook, and Schmitt 2001; G. Bernardi and Crane 1999; G. Bernardi et al. 2002; McCafferty et al. 2002, Roberts et al in prep). In addition, this study also found low gene flow (Fst =0.19) between the Indian Ocean and the Pacific *D. trimaculatu*s clade (Table 3) – which is consistent with previous studies: [Zanzibar-Japan Fst=0.23 (Giacomo Bernardi, Holbrook, and Schmitt 2001); Pacific-Indian Fst=0.25 (Salas et al. 2020)]. The admixture analysis (Figure 2) also shows clear distinction between samples from Oman (Indian Ocean) and all other samples from the Pacific Ocean. Salas et al (2020) showed that the boundary between the Indian Ocean and Pacific Ocean populations of *D. trimaculatus* occurs in the eastern Indian Ocean where a hybridization zone occurs around Cocos-Keeling and Christmas Islands. This region between the Australian continent and the

Indo-Malay shelf is narrow and shallow and a known biogeographic boundary region for many marine sister species (Gaither and Rocha 2013). Morphological differences, including clear dorsal fin in the Pacific Ocean, and a dark dorsal fin in the Indian Ocean, also fully partition the two populations (Figure 1).

The other clear genetic separation was at the opposite end of the current *D. trimaculatus* distribution. We found that samples from French Polynesia in Mo'orea exhibit less gene flow (Fst=0.22) with the Pacific *D. trimaculatus* clade than the Indian Ocean clade. The admixture plot supports previous findings using microsatellite data showing a clear isolation of all Mo'orea individuals (Leray et al. 2010). Leray et al suggest that the relatively short PLD (23-30 days) potentially paired with differential survival of larvae restricts gene flow enough to explain the observed genetic divergence. In the tropical marine system, the highest rates of endemism are found in peripheral populations of the Indo-Malay-Philippines Archipelago and it is likely that this branch of *D. trimaculatus* constitutes another endemic species of the Society Islands (Bellwood and Wainwright 2002; Connolly, Bellwood, and Hughes 2003; Bowen et al. 2016; Briggs and Bowen 2012). It would join other peripheral endemic species in this complex found at other known endemic hotspots, *D. strasburgi* (Marquesas) and *D. albisella* (Hawai'ian archipelago), in this study.

Pacific-wide population genomics

We find that, interestingly, not all peripheral populations of the *D. trimaculatus* complex included in this study are as divergent as the French Polynesian population. For example, *D. auripinnis* (Randall and Randall 2001) is the most recently described species of the group, characterized by varying levels of orange-yellow to the belly and pelvic, anal, and caudal fins, and flecks of yellow along dorsal spine margins. Its published range includes the Line Islands (represented by Kingman Reef in this study), Phoenix Islands (represented by Baker Atoll in this study), and the Northern Cook Islands. Leray et al (2010) found genetic evidence of introgression to the south in Fiji. In this study, we find strong evidence for genetic separation (black genotyopes, Figure 3) but also evidence of introgression into the greater Pacific clade (Fst = 0.005). Signals of *D. auripinnis* are prominent in Fiji which support findings by Leray et al (2010) but also in Ulithi over 5000km to the west, and also present at low levels as far as the Western Pacific in southern Japan. The admixture analysis (Fig 2) also shows that admixture seems to be directional, likely with *D. auripinnis* in Baker and Kingman Reef acting as a source of genetic diversity into the Pacific clade while remaining largely impervious to immigration. The west bound Equatorial Currents in which the Phoenix and Line Islands is likely to play a role in both isolating the Phoenix and Line Islands from outside genetic influence while aiding in the export of propagules to the greater western Pacific.

The Pacific clade, which, in our study is represented by individuals from the Philippines, Ryukyus, Ulithi, and Fiji, is characterized by four main genetic lineages based on admixture analysis of k=7. Fiji, seems to represent a sink population being

the only population to show elements of all four genetic clusters. Its main component is shared with *D. auripinnis* in black, along with a minor but uniform proportion of Mo'orea polymorphisms (green). Interestingly, despite being geographically more distant from the Philippines (pink) than Ulithi which is only roughly a quarter of the distance to the Philippines, most samples from Fiji show some proportion of admixture with the Philippine lineage. Only two Ulithian fish on the other hand, share any of the Philippine genomic signature and is otherwise relatively uniformly characterized by *D. auripinnis* and the Pacific teal lineage.

In Chapter 1, the mitochondrial dloop from the Ryukyu archipelago largely clustered into a sub clade of the greater Pacific clade. Data based on low coverage whole genome sequencing in this study reveal several interesting characteristics. First, unlike the other peripheral populations we examined here, the Ryukyu population is not completely isolated (Table 2, Figure 2). Both gene flow estimations and admixture analysis show that there is clear exchange with the rest of the Pacific clade and to some extent *D. auripinnis*, as discussed earlier. Most of the Ryukyu population seems to be characterized primarily by a combination of two lineages: one clearly from the Philippines seen in pink in Figure 2 and the other depicted by the teal bars seen across the Ryukyus and also prominent in Ulithi Atoll located just south of the Marianas Islands. Interestingly, again, gene flow seems to be strongly directional between the Philippines and Ryukyu Archipelago. While again, it is difficult to fully identify specific mechanisms, one important factor is likely to be a strong current, the Kuroshio Current, that drives both warm tropical water and propagules from the

Philippines northward along the Ryukyu Archipelago. As discussed in Chapter 1, this system is one in which both high gene flow (aided by the Kuroshio Current) and selection (related to the latitudinal gradient of the subtropics) are likely opposing forces at play.

Admixture results of Philippine and Japanese lineages show a distinct genetic cline. Philippine samples belong to a single genetic cluster (pink) except one individual, while 16% of Japanese samples belong to another single lineage (teal). The presence of a genetic cline suggests ongoing gene flow and intermixing of the two lineages within the Ryukyu archipelago. Interestingly, while the cline comprises individuals with all potential fractions of genetic clusters, individuals below 25% of either Philippine (pink) or Japan (teal) genetic components, are missing. This might be due to selection against such individuals, or sterility in their potential parents. Though more analysis is needed to tease apart the mechanism, there are several possible mechanisms by which this result might arise: 1) an old lineage in the Ryukyus followed by a Philippines range expansion and subsequent introgression 2) a continual input of Philippine propagules that are selected against leaving a Ryukyu adapted linages.

Modes of speciation in the D. trimaculatus species complex

Speciation modes have been described as a continuum between allopatric speciation, where populations diverge as a result of geographic separation, to sympatric speciation which occurs despite geographic overlap. Between those

extremes, parapatric and peripatric speciation may also occur where geographic isolation is incomplete. In the *D trimaculatus* complex, the edge populations of *D. albisella* in the Hawai'ian Islands, *D. strasburgi* in the Marquesas and the divergent populations of *D. trimaculatus* in the Society Islands and again in the Indian Ocean, likely represent examples of allopatric speciation where some combination of currents, distance, ecology, and biology have isolated these populations and favored their divergence. *Dascyllus auripinnis*, on the other hand, may represent a peripatric scenario of speciation where this restricted edge population divergence occurred during a temporary period(s) of isolation and is currently acting as a source of genetic diversity to the greater Pacific *D. trimaculatus* population. Finally, the edge population that is here represented by Japanese samples, is potentially an example of incipient speciation in the face of gene flow, where some individuals are already sorted (fully teal individuals), while others are still carrying evidence of gene flow.

Conclusion

In this study, using whole genome sequencing, we find clear and varied patterns of population structure in the *Dascyllus trimaculatus* species complex. We recommend additional taxonomic designations for the Pacifc *D. trimaculatus* as well as those in the Society Islands. Some taxonomic names might be available such as *D. niger,* described in 1847, and *D. nuchalis*, described in 1830, both from Indonesia, and both currently synonymized with *D. trimaculatus.* Our findings also make clear that there is much to be understood about even the basic population structure of this

species complex and the speciation mechanisms. For example, how far do the Philippine and Ryukyu genotypes extend? Is the cline of the Philippine lineage in Southern Japan a result of selection, incomplete lineage sorting, or perhaps another case of peripatry and subsequent lineage sorting? There are clearly many more to questions to be investigated to better understand how this species complex came to be what it is today, and what we may expect in the future.

Acknowledgements

Funding: Dissertation Year/Quarter Fellowship (Division of PBSci), Visiting Scholar Fellowship-PI Mike Berumen (Red Sea Research Center, KAUST), Lewis and Clark Fund for Exploration, NSF-JSPS East Asia Pacific Summer Institute Fellowship, STARS ReEntry scholarships, EEB summer funding. **Permits and collection permissions**: Jayvee Sacco, Angela Quieros, Sea Institute, John Magul Rulmal Jr, James Reimer, Shiori, Iori, Takuma Fuji **Sampling help or source:** Milo Tasopulu & Nickson Talgusey (Ulithi), Remy (Philippines, Hawai'i), Bernardi Lab (Hawai'i), Iori Kawamura(Okinawa), Toru Naruse (Iriomote), Giacomo Bernardi (Hawai'i, Mo'orea, Marquesas, Fiji, Baker Atoll Kingman Reef, Oman) **Lab:** Anna Worth, Rachel Meyer, Evan Pepper **Analysis:** Liam Zarri, Brian Simison

References:

Bellwood, David R, and Peter C Wainwright. 2002. “Chapter 1: The History and Biogeography of Fishes on Coral Reels.” In *Coral Reef Fishes*, 28. Elsevier Science.

Bernardi, G., and N. L. Crane. 1999. “Molecular Phylogeny of the Humbug Damselfishes Inferred from MtDNA Sequences.” *Journal of Fish Biology* 54 (6): 1210–17. https://doi.org/10.1111/j.1095-8649.1999.tb02049.x.

Bernardi, G., S. J. Holbrook, R. J. Schmitt, N. L. Crane, and E. DeMartini. 2002. “Species Boundaries, Populations and Colour Morphs in the Coral Reef Three-Spot Damselfish (Dascyllus Trimaculatus) Species Complex.” *Proceedings of the Royal Society B: Biological Sciences* 269 (1491): 599–605. https://doi.org/10.1098/rspb.2001.1922.

Bernardi, Giacomo, Sally J. Holbrook, and Russell J. Schmitt. 2001. “Gene Flow at Three Spatial Scales in a Coral Reef Fish, the Three-Spot Dascyllus, Dascyllus Trimaculatus.” *Marine Biology* 138: 457–65.

Bhatia, Gaurav, Nick Patterson, Sriram Sankararaman, and Alkes L. Price. 2013. “Estimating and Interpreting F_{ST} : The Impact of Rare Variants.” *Genome Research* 23 (9): 1514–21. https://doi.org/10.1101/gr.154831.113.

Bowen, Brian W., Michelle R. Gaither, Joseph D. DiBattista, Matthew Iacchei, Kimberly R. Andrews, W. Stewart Grant, Robert J. Toonen, and John C. Briggs. 2016. “Comparative Phylogeography of the Ocean Planet.” *Proceedings of the National Academy of Sciences* 113 (29): 7962–69. https://doi.org/10.1073/pnas.1602404113.

Briggs, John C., and Brian W. Bowen. 2012. “A Realignment of Marine Biogeographic Provinces with Particular Reference to Fish Distributions.” *Journal of Biogeography* 39 (1): 12–30. https://doi.org/10.1111/j.1365-2699.2011.02613.x.

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. “Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor.” *Bioinformatics* 34 (17): i884–90. https://doi.org/10.1093/bioinformatics/bty560.

Connolly, Sean R., David R. Bellwood, and Terence P. Hughes. 2003. “Indo-Pacific Biodiversity of Coral Reefs: Deviations from a Mid-Domain Model.” *Ecology* 84 (8): 2178–90. https://doi.org/10.1890/02-0254.

Danecek, Petr, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, et al. 2021. “Twelve Years of SAMtools and

BCFtools." *GigaScience* 10 (2): giab008. https://doi.org/10.1093/gigascience/giab008.

Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." *Bioinformatics* 32 (19): 3047–48. https://doi.org/10.1093/bioinformatics/btw354.

Gaither, Michelle R., and Luiz A. Rocha. 2013. "Origins of Species Richness in the Indo-Malay-Philippine Biodiversity Hotspot: Evidence for the Centre of Overlap Hypothesis." *Journal of Biogeography* 40 (9): 1638–48. https://doi.org/10.1111/jbi.12126.

Korneliussen, Thorfinn Sand, Anders Albrechtsen, and Rasmus Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." *BMC Bioinformatics* 15 (1): 356. https://doi.org/10.1186/s12859-014-0356-4.

Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Leray, Matthieu, Ricardo Beldade, Sally J. Holbrook, Russell J. Schmitt, Serge Planes, and Giacomo Bernardi. 2010. "Allopatric Divergence and Speciation in Coral Reef Fish: The Three-Spot Dascyllus, Dascyllus Trimaculatus, Species Complex." *Evolution* 64 (5): 1218–30. https://doi.org/10.1111/j.1558-5646.2009.00917.x.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

McCafferty, S., E. Bermingham, B. Quenouille, S. Planes, G. Hoelzer, and K. Asoh. 2002. "Historical Biogeography and Molecular Systematics of the Indo-Pacific Genus Dascyllus (Teleostei: Pomacentridae)." *Molecular Ecology* 11 (8): 1377–92. https://doi.org/10.1046/j.1365-294X.2002.01533.x.

Randall, John E, and Helen A Randall. 2001. "*Dascyllus auripinnis*, a New Pomacentrid Fish from Atolls of the Central Pacific Ocean." *Zoological Studies*, 7.
Salas, Eva M., Jean-Paul A. Hobbs, Moisés A. Bernal, W. Brian Simison, Michael L. Berumen, Giacomo Bernardi, and Luiz A. Rocha. 2020. "Distinct Patterns of Hybridization across a Suture Zone in a Coral Reef Fish (*Dascyllus trimaculatus*)." *Ecology and Evolution* 10 (6): 2813–37. https://doi.org/10.1002/ece3.6068.

Skotte, Line, Thorfinn Sand Korneliussen, and Anders Albrechtsen. 2013. "Estimating Individual Admixture Proportions from Next Generation Sequencing Data." *Genetics* 195 (3): 693–702. https://doi.org/10.1534/genetics.113.154138.
Wellington, G. M., and B. C. Victor. 1989. "Planktonic Larval Duration of One Hundred Species of Pacific and Atlantic Damselfishes (Pomacentridae)." *Marine Biology* 101 (4): 557–67. https://doi.org/10.1007/BF00541659.

Ch3 - Figures and Tables

Figure 1. D. trimaculatus species complex distribution and sampling locations. The three-spot damselfish species complex with previously known distributions and sampling locations for the current study is shown on the map adapted from Leray et al. (2009) and Salas et al. (2020). Corresponding illustrations of typical morphs/species for the regions/species are pictured. The background color of the fish illustrations corresponds to its distribution in the same color. The Indian Ocean clade is shown in purple, the greater Pacific clade in blue, *D. albisella* in green, *D. strasburgi* in red, *D. auripinnis* in peach, and *D. trimaculatus* Society Islands population in yellow.

Figure 1.

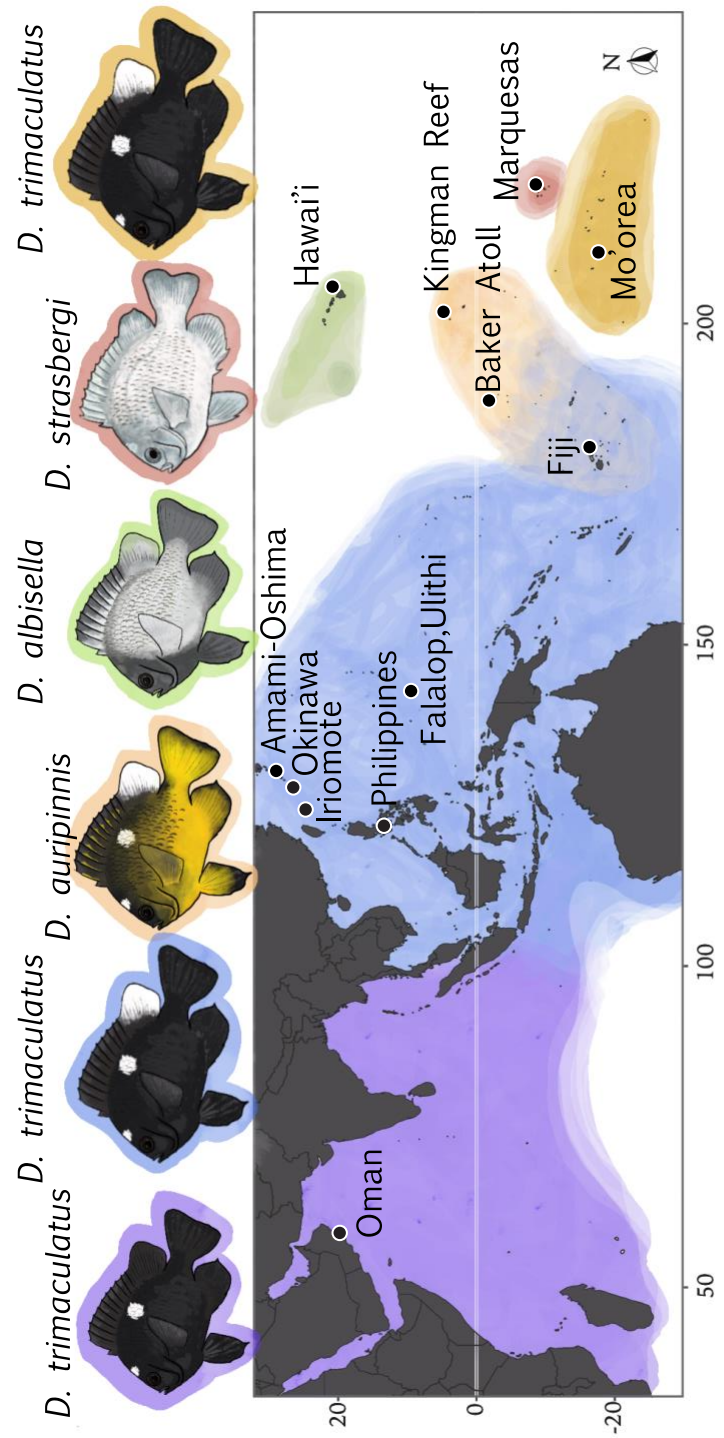

D. trimaculatus
D. trimaculatus
D. auripinnis
D. albisella
D. strasbergi
D. trimaculatus
Oman
Amami-Oshima
Okinawa
Iriomote
Philippines
Falalop,Ulithi
Hawai'i
Kingman Reef
Baker Atoll
Fiji
Marquesas
Mo'orea
N
20
0
-20
50
100
150
200

Figure 2. Admixture plot of D. trimaculatus species complex. This admixture bar plot k=7 (angsd v 0.939 NGSadmix) shows estimated proportions of individual admixture using genotype likelihoods calculated for 132 individuals ordered by population (Oman, n=11; Hawai'i, n=9; Marquesas, n=14; Mo'orea, n=11; Philippines, n=10; Iriomote, n=10; Okinawa n=11; Amami-Oshima n=11; Ulithi n=11; Fiji n=11; Kingman Reef n=10; Baker Atoll n=10).

Figure 2.

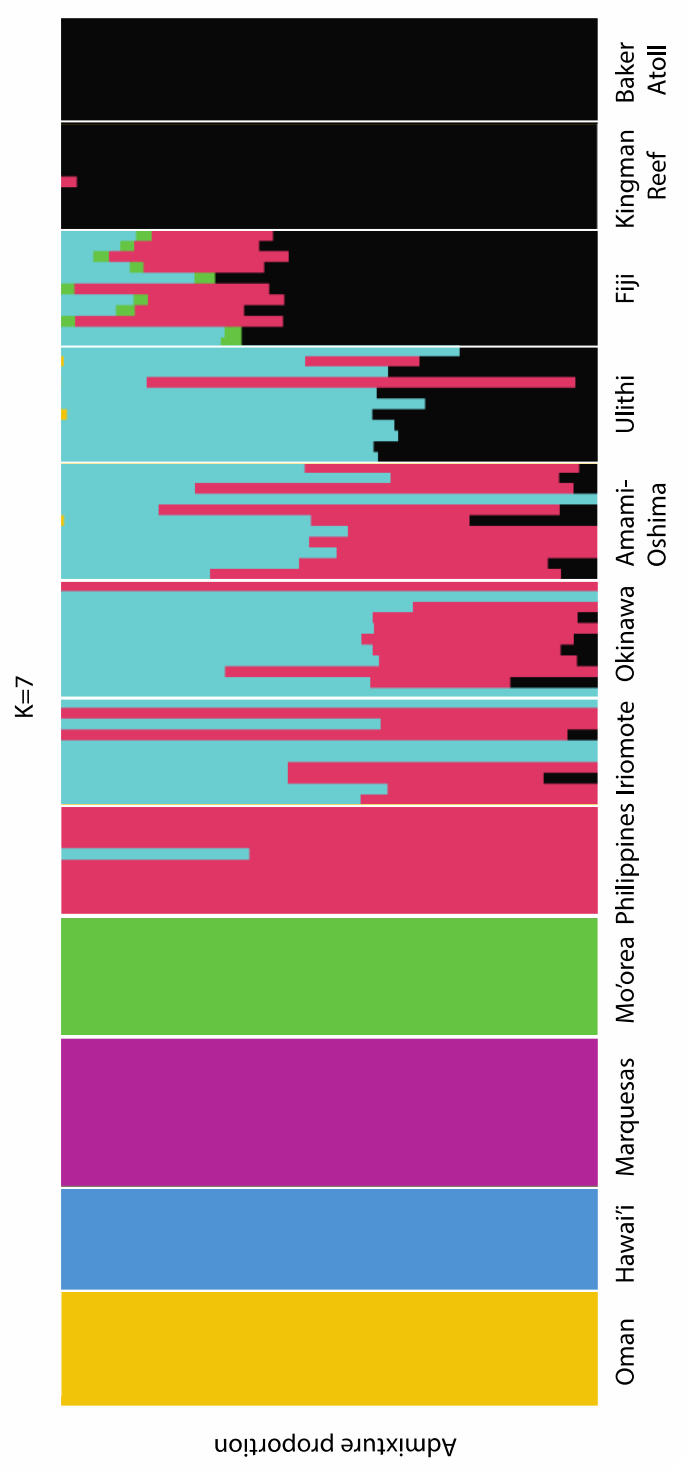

K=7
Admixture proportion
Oman
Hawai'i
Marquesas
Mo'orea
Philippines
Iriomote
Okinawa
Amami-Oshima
Ulithi
Fiji
Kingman Reef
Baker Atoll

Figure 3. Subset of k=7 admixture bar plot. A subset of Figure 2 plot showing only Philippines and Ryukyu Archipelago individuals and ordered by proportion of Philippine admixture (pink).

Figure 3.

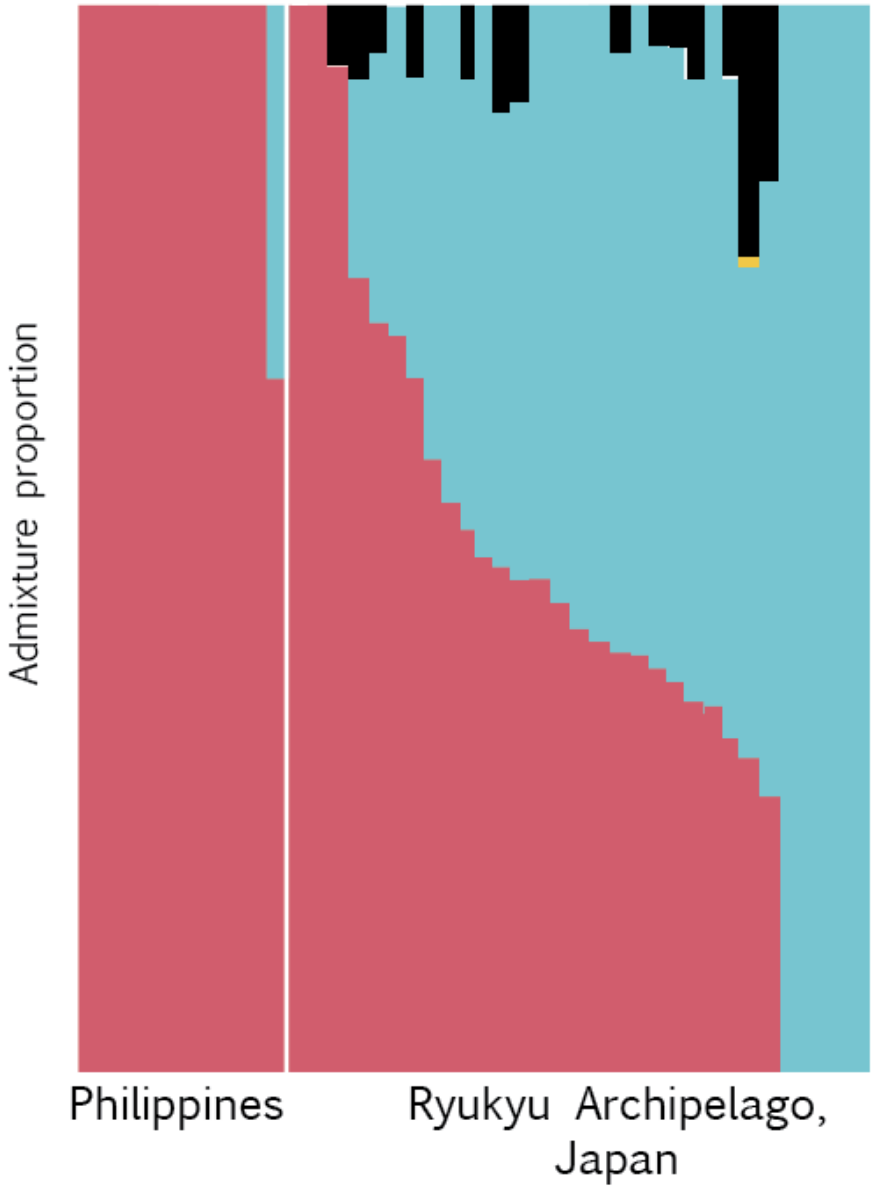

Admixture proportion
Philippines
Ryukyu Archipelago, Japan

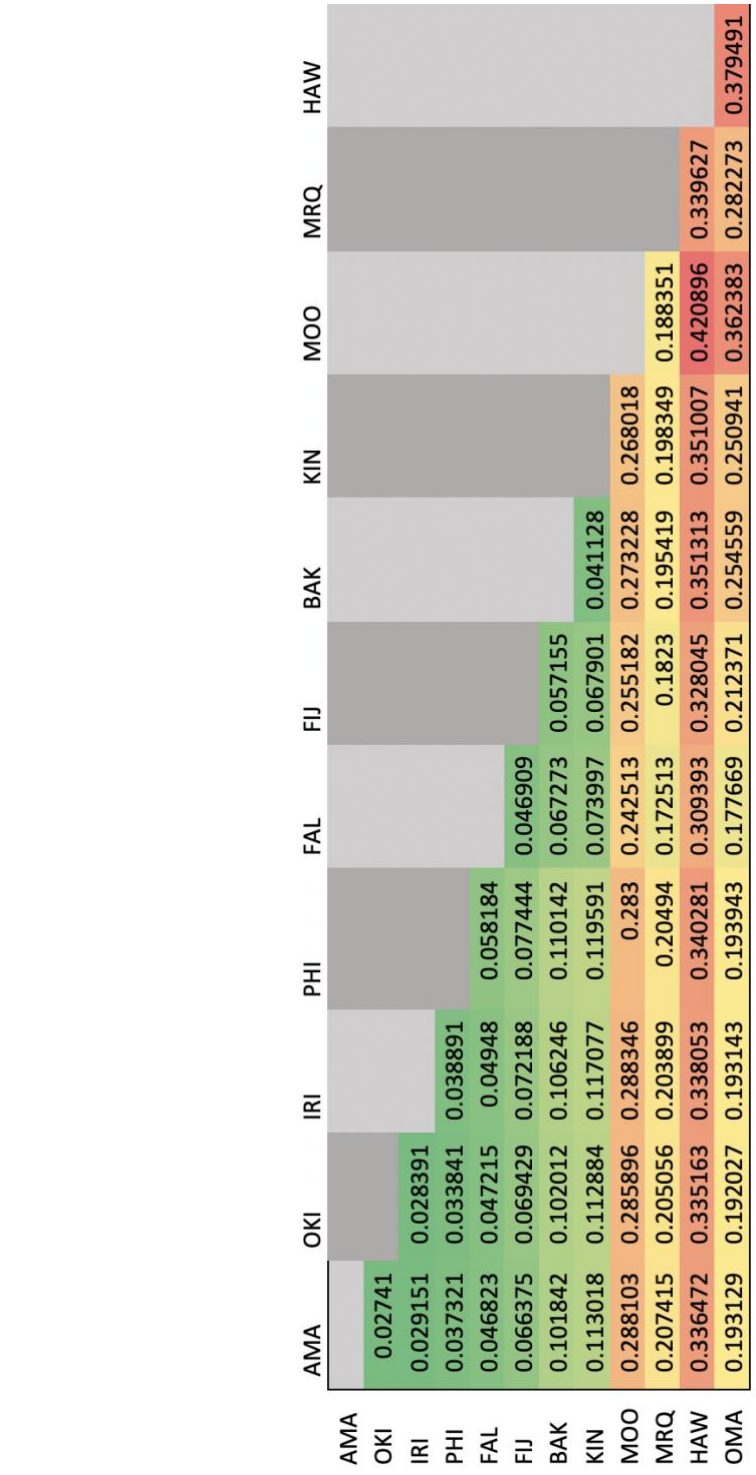
Table 1. Sample table with published species assignment, location of collection, and number of individuals in the *D. trimaculatus* species complex used in the study.

Table 1.

Species assignment	Collection location	n
Dascyllus albisella	Hawai'i, Hawai'ian Islands	9
Dascyllus strasburgi	Marquesas Islands	14
Dascyllus auripinnis	Kingman Reef	10
Dascyllus auripinnis	Baker Atoll	10
Dascyllus trimaculatus	Oman, Indian Ocean	11
Dascyllus trimaculatus	Mo'orea, Society Islands	11
Dascyllus trimaculatus	Amami-Oshima, Ryukyu Archipelago	11
Dascyllus trimaculatus	Okinawa, Ryukyu Archipelago	11
Dascyllus trimaculatus	Iriomote, Ryukyus Ryukyu Archipelago	10
Dascyllus trimaculatus	Verde Island, Passage, Philippines	10
Dascyllus trimaculatus	Ulithi, Micronesia	11
Dascyllus trimaculatus	Fiji	11

Table 2. Weighted pairwise Fst. Pairwise weighted Fst are reported between sites below the grey diagonal. The color spectrum represents levels of geneflow – darker greens represent higher gene flow to reds which represent low to lack of gene flow based on Fst values. Population codes and published species designations are as follows: *D. trimaculatus* (AMA = Amami-Oshima, OKI=Okinawa, IRI-Iriomote, PHI=Philippines, FAL=Falalop, FIJ=Fiji), *D. auripinnis* (BAK=Baker Atoll, KIN=Kingman Reef), *D. trimaculatus* (MOO=Mo'orea), *D. strasburgi* (MRQ=Marquesas), *D. albisella* (HAW=Hawai'i), *D. trimaculatus* (OMA=Oman).

Table 2.



	AMA	OKI	IRI	PHI	FAL	FIJ	BAK	KIN	MOO	MRQ	HAW
AMA											
OKI	0.02741										
IRI	0.029151	0.028391									
PHI	0.037321	0.033841	0.038891								
FAL	0.046823	0.047215	0.04948	0.058184							
FIJ	0.066375	0.069429	0.072188	0.077444	0.046909						
BAK	0.101842	0.102012	0.106246	0.110142	0.067273	0.057155					
KIN	0.113018	0.112884	0.117077	0.119591	0.073997	0.067901	0.041128				
MOO	0.288103	0.285896	0.288346	0.283	0.242513	0.255182	0.273228	0.268018			
MRQ	0.207415	0.205056	0.203899	0.20494	0.172513	0.1823	0.195419	0.198349	0.188351		
HAW	0.336472	0.335163	0.338053	0.340281	0.309393	0.328045	0.351313	0.351007	0.420896	0.339627	
OMA	0.193129	0.192027	0.193143	0.193943	0.177669	0.212371	0.254559	0.250941	0.362383	0.282273	0.379491

Table 3. Weighted pairwise Fst by species and major Pacific clades. Pairwise Fst are reported between populations grouped by species or major clades. Pairwise weighted Fst are reported between sites below the grey diagonal. The color spectrum represents levels of geneflow – greens represent higher gene flow, yellows intermediate, and reds represent low to lack of gene flow based on Fst values. Group codes are defined as follows: PAC is *D. trimaculatus* which includes AMA = Amami-Oshima, OKI=Okinawa, IRI-Iriomote, PHI=Philippines, FAL=Falalop, FIJ=Fiji; DAU is *D. auripinnis* which includes BAK=Baker Atoll, KIN=Kingman Reef; MOO is *D. trimaculatus* from the Mo'orea/ Society Islands clade, DST is *D. strasburgi* from Marquesas, DAL is *D. albisella* from Hawai'i; IND is *D. trimaculatus* from Oman in the Indian Ocean.

Table 3.

	PAC	DAU	MOO	DST	DAL
DAU	0.005836				
MOO	0.222942	0.228085			
DST	0.170429	0.173031	0.188351		
DAL	0.306278	0.312101	0.420896	0.339627	
IND	0.194267	0.23644	0.362383	0.282273	0.379491

Supplementary Protocols

SP1. DNA extraction protocol (2 pages)

Chloroform DNA extraction Protocol_v1 M. Roberts 11.2020

Read through the entire protocol + notes on right before starting!

Materials

Tissue cutting tool (ex: razor blade, scissors etc)
Sterilizing materials (ex: flame, EtOH, water)
Tubes 1.5-2ml (2x per sample)
Lysis Buffer
20%SDS
Proteinase K
Chloroform
6M NaCl
RNAase
UltraPure H20

Starting

* Protocol requires an overnight digestion.
* Most steps are best done with a manual pipet.
* This protocol requires 2 x 1.5ml tubes not including the tube the tissue samples come in.

Set oven to 56C

Step 1

Add small piece of each sample to 1.5ml tube, EtOH should be dried off

Step 2

Make a master mix – invert to mix well

613 µl Lysis buffer
30 µl 20% SDS x # samples + 2-3
7 µl Proteinase K

Step 3

Add **650 µl of master mix** to each tube and mix well.
Incubate in oven for ~3 - 12 hours/overnight.

Step 4

After digestion, pull from oven and add **375 µl of 6M NaCl** and slowly invert tubes to mix. Allow particulate to settle out of solution for ~30 min. Mix every ten by inverting tube

Step 5

Centrifuge at 14,000rpms for 30 min
Label new tubes

Step 6

Remove 700 µl of supernatant and put in new tube

Step 7

Add **700µl chloroform** to each sample, mix by turning tubes slowly
Centrifuge at 12,000rpms for 10min
Set oven to 37C

Step 8

Remove top clear portion of supernatant and put in clean labeled tube. (~600µl)

* put chloroform in cap of bottle, pull chloroform into pipet tip several times to coat before adding to first sample

*Be careful not to aspirate particles from the boundary. If you do push back out into original tube and only save clear liquid. A wider tip helps with this so consider cutting tips with a razor.

Step 9
Add **1 µl RNAase A** (10mg/ml) (RNA is stored in freezer by Giacomo's station).
Incubate at 37C for 10 minutes
Step 10
Add **700 µl Isopropanol** and mix until you can see DNA
Centrifuge for 30 minutes at 14,000 RPMs *have cap hinge outward
Step 11
Remove supernatant by pouring down the drain, careful not to lose the pellet. Turn tube upside down onto a towel to remove liquid from rim and leave tube open
Use P20 to remove excess liquid before drying again careful not to remove pellet
Step 14
Put samples in speed vac and run program 1 until dry (usually twice for a total of 4 minutes, but use judgement.
Once dry **add 150 µl H2O** and store in fridge if using soon, store in freezer if waiting to prep library, but allow DNA to rehydrate first.

You're done. Congrats.

Lysis buffer:
10 mM Tris
400mM NaCl
2mM Na2 EDTA

SDS:
20% solutions
2 grams of SDS powder (stored in dry chemical cabinet) to 8g H2O

SP2 – Tn5 whole genome library preparation protocol (3 pages)

Tn5 Library Prep Protocol_v3

M. Roberts 05.2021

**This protocol is adapted from previous version from E. Pepper and A. Worth.
Read through the entire protocol +notes on right before starting!

I. Annealing Oligos and Loading Tn5

1. Thaw out 100 µM aliquots of the oligos

*Do not thaw the Tn5 - leave in freezer

Tn5ME – R: (reverse complement of A and B) has phosphate
Tn5ME – A: i5 forward
Tn5ME – B: i7 reverse

Label 2 PCR tubes "AR" and "BR"

2. Add equal volumes of oligo "R" to both tubes, and equal volumes of oligo "A" or oligo "B" to their respective tubes.

*2µl minimum to reduce pipetting error and to account for evaporation (this makes enough for 32 libraries
– or-
Calculate an extra 4 libraries to account for pipettor error, shits-and-giggles, etc.

0.0715µl oligo/rxn
-OR-
2ul

Mix via pipetting up and down 10 times then spin down.

3. Anneal the oligos in thermocycler (~8min).

95°C for 5 minutes
95°C to 25°C at -1 °C/second
25°C infinite hold

Label two more PCR tubes "Tn5 + AR" and "Tn5 + BR".

Spin the PCR tubes briefly to bring any condensation liquid down to the bottom of the tube

4. Add Tn5 (viscous) and respective annealed oligos to both PCR tubes labeled "Tn5 + AR" (or BR)

*Since Tn5 is an enzyme, keep it in the freezer until you actually need it and to put it back in the freezer right when you are done with it.
this is, in cool kid lingo, "charging the Tn5" with the oligos. Tn5
*Calculate for an extra 4-5 samples, once this mixes with TAPS later on, some is lost on outside of pipette etc.

Tn5:	1µl/rxn		=____	Tn5:	1µl/rxn		=____
	+	x____			+	x____	
AR:	0.143µl/rxn		=____	BR:	0.143µl/rxn		=____

Pipette mix well then spin down.

5. Incubate for 1 hour at room temperature to charge the Tn5 (Tn5*).
Start Time_____

*pipette mix well

*If you need longer (snack time?!) incubate at room temperature for one hour then place if the refrigerator. (Charged Tn5 will stay good for use in the fridge for several hours after a one hour room temperature activation.)

If you haven't already, you have shit to do my friend:
1. Bead clean DNA to get rid of pieces <1000bp
2. Qubit (gotta know your DNA concentrations)
3. Dilute or calculate #µl needed to get 10-20ng of DNA for the tagmentation steps
4. Label next couple sets of tiny tubes:
   a. Label tubes – 1 per sample for tagmentation step
   b. Label tubes – 1 per sample for indexing steps

6. Combine "Tn5 + AR" and "Tn5 + BR" into a single tube, pipet mix.

**This protocol is adapted from previous version from E. Pepper and A. Worth.
Read through the entire protocol +notes on right before starting!

II.DNA Tagmentation *PREPARE ON ICE

Tagmentation requires 10-20ng of HMW DNA

1. Prepare Master Mix:

Charged Tn5 (Tn5*)…….. 2µl/rxn x_____ _____ul of Tn5*
5x TAPS-PEG8000:……..4µl/rxn _____ul of TAPS

>If your DNA is relatively consistent in concentrations and are including water into your master mix:

Ultra Pure H20:……13µl/rxn (1µl DNA) x_____ _____ul of H2o
~OR~ ……12µl/rxn (2µl DNA) x_____ _____ul of H2o
…… **Etc. (DNA +H20 = 14ul vol)**

~OR~

>If not:
This is when you add the right amount of water and DNA to your new tubes
DNA +H20 = 14ul vol

2. For samples that require:

If you made a MM with water included then in a new tube per sample:
1µl of DNA, add 19µl of MM
~OR~
2µl of DNA, add 18µl of MM etc

If you made a MM **without water** included then in a new tube per sample:
DNA +H20 = 14ul
MM (Tn5* + 5x TAPS-PEG8000). = 6ul

Total reaction volume (MM + DNA should to be 20µl)
Mix via pipetting and spin down

3. Incubate the reaction in the thermocycler at 55°C for 8 minutes.

Prepare to quickly perform step 4.Get 0.2% SDS ready in a sterilized boat/tray, and your mulit channel set and ready to go.

If you haven't already, label those indexing tubes!

4. Kill the Tn5 reaction by quickly adding 5 µl of 0.2% SDS and mix via pipetting up and down. Let this incubate at room temperature for at least 7 minutes

*Again, **PREPARE ON ICE**- Tn5 becomes active at room temperature

*Calculate for an extra 4-5 samples, once this mixes with TAPS later on, some is lost on outside of pipette etc.

* Total rxn vol needs to be 20µl so H20 vol needs to be adjusted. Sometimes it makes sense to make 2 + Master Mixes to take this into account othertimes it makes sense for the MM to be only the Tn5* and TAPS and to add the water and DNA individually. In this last case just add 6ul of the MM to each tube

*5x TAPS-PEG8000 is VERY VISCOUS pipette slowly and pipetet mix well (hard to do with an e-pippette

*The **incubation time is important**:
*The incubation time and the volume of Tn5 added can be modulated, to optimize the tagging of a given sample. If your DNA is highly degraded, consider using less Tn5 in the tagmentation - loading too much will fragment your DNA too much, generating primarily short reads. Always keep in mind the quality of your sample going into this reaction.

*SDS needs to be pipette mixed in well, it's also viscous
* This can take time so do your more degraded DNA first.

**This protocol is adapted from previous version from E. Pepper and A. Worth.
Read through the entire protocol +notes on right before starting!

III. PCR Indexing and Amplification*PREPARE ON ICE

*Thaw/keep your index primers on ice!
*You will indexing your samples in this step

1. Prepare a master mix (MM) (per rxn)

UltraPure H20	11.75µl/rxn		=_____ul H2o
5xKAPA HiFi Buffer	5µl/rxn	x_____	=_____ulKAPABuffer
dNTPs	0.75µl/rxn		=_____uldNTPs
HiFi Polymerase	0.5µl/rxn		=_____ulPolymerase

Total vol of MM per rxn……..18µl/rxn

2. Add seperately to each tube:

Combined 10uM i5 & 10uM i7 index primers ……2µl
Tn5 Product ……..5µl
Master Mix ……..18µl

Total vol(MM+indexprimers+DNA)…….25µl

Pipet mix and spin down

3. Run PCR protocol in thermocycler: (~40-45min)

72°C for 5 minutes
95°C for 3 minutes
98°C for 20 seconds
65°C for 15 seconds **x 12 cycles**
72°C for 30 seconds
72°C for 5 minutes
Hold at 4°C ∞

Take the SPRI beads out to get to room temp for bead cleaning step.
Label final library tubes, Qubit tubes, prep Qubit.

4. Qubit libraries to make sure you've obtained sufficient amplification.

5. If all Qubit values are reflective of successful library generation (30--60 ng/µl), pool all of the technical replicate PCR reactions together. You will end up with a single tube per sample.

6. Bead clean the pooled libraries using 1x (1:1)
Should have 23µl of library - so use 23µl SPRI beads (2µl went to Qubit)
Elute with 30-40µl

7. Qubit libraries to determine concentration of purified DNA library

8. Store libraries in -20°C

*Again, **PREPARE ON ICE**- Non-Hotstart KAPA HiFi Polymerase is active at room temperature.

*Keep everything on ice!
*Factor in pipette error into master mix calculations

*Multi-channel is a *DREAM* here to add the Tn5 product and indexes.

* Bolded steps are the cycled steps

* Remember, if 10 ng of DNA went into the tagmentation reaction that was brought to 25µl, and only 5µl of that was put into the PCR reaction, then your initial DNA concentration of the 25µl PCR reaction is ~0.08 ng/µl. Good amplification after 12 cycles should yield anywhere from 30 -- 60 ng/µl.

* Getting rid of small fragments, unused ingredients, and other junk.

***after freezing and thawing, Qubit should be redone to figure out [ng/µl] prior to pooling**.

Conclusion

This dissertation builds upon previous research of this complex which had made it a model system in its own right and adds genomic resources to ensure that we continue to further our understanding of this important model system in this genomic age. There are now resources for endless directions and questions to follow regarding this species complex and evolutionary processes. For example, Chapter 1 makes clear that interesting and unique processes are occurring in unexamined corners of what is thought to be a well-mapped out species. What will other regions find? The chromosome-level genome presented in Chapter 2 highlights the question of the role that inter-chromosomal connections (prevalent in this genus) may play in speciation processes and perhaps even the formation of this complex. We can follow up on Chapter 3 findings of the cline of southern genotypes in the northern most reefs of the Pacific in the Ryukyu archipelago to delve into the mechanisms and intricacies of selection and local adaptation.

Understanding mechanisms of dispersal, the role of selection, and the capabilities of local adaptation in a species with relatively restricted gene flow are some of the exciting questions that this project has uncovered, thus placing *Dascyllus trimaculatus* as an ideal model species to address fundamental questions in the ecology and evolution of coral reef fishes.