Tale of the Two-Spot: Three Investigations into the Evolutionary History of the Octopus

by

Gabrielle Colleen Coffing

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Biology

Dissertation Committee:

Cris Niell, Chair

Andrew Kern, Advisor

Adam Miller, Core Member

Mara Lawniczak, Core Member

Peter Ralph, Institutional Representative

University of Oregon

Summer 2025

DISSERTATION ABSTRACT

Gabrielle Colleen Coffing

Doctor of Philosophy in Biology

Title: Tale of the Two-Spot: Three Investigations into the Evolutionary History of the Octopus

Cephalopods (octopus, squid, cuttlefish, and nautilus) are a diverse class of species that are famous for their morphological innovations including their unique body plans with sucker-bearing arms, complex nervous systems, and camouflaging abilities. By examining sequencing datasets of cephalopods, we can learn about how their collection of novel traits has evolved. In this dissertation, I explore the evolutionary history of the California two-spot octopus (*Octopus bimaculoides*) in three different projects. In Chapter II, I assemble a high-quality genome assembly and annotation which we harnessed to determine the transcriptional cell types of the octopus visual system. In Chapter III, I discovered that octopus have a sex chromosome and show that it is shared across all extant cephalopods. Finally, in Chapter IV, I conducted a descriptive population genetics study on *O. bimaculoides* and its sister species, *O. bimaculatus*, to show that their genetic diversity and divergence are consistent with their life history strategies. Together, my dissertation explores the evolutionary past of the octopus.

This dissertation includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Gabrielle Colleen Coffing

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
DePauw University, Greencastle, IN

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2025, University of Oregon
Bachelor of Arts, Cellular and Molecular Biology & Computer Science, 2019, DePauw University

AREAS OF SPECIAL INTEREST:

Evolutionary Biology
Genomics
Bioinformatics

GRANTS, AWARDS, AND HONORS:

Donald E. Wimber Award, University of Oregon, 2023

Best Poster Award, Cephalopod International Advisory Council, 2022

Graduate Research Fellowship, National Science Foundation, 2021- 2025

Genetics Training Grant, University of Oregon, 2020-2021

Goldwater Scholar, Barry Goldwater Scholarship and Excellence in Education Foundation, 2018-2019

PUBLICATIONS:

Coffing, G.C., Tittes, S., Small, S.T., Songco-Casey, J.O., Piscopo, D.M., Pungor, J.R., Miller, A.C., Niell, C.M., and Kern, A.D. (2025). Cephalopod sex determination and its ancient evolutionary origin. Current Biology *35* 931-939.e4. https://doi.org/10.1016/j.cub.2025.01.005.

Baden, T., Briseño, J., **Coffing, G.C.,** Cohen-Bodénès, S., Courtney, A., Dickerson, D., ... & Albertin, C. B. (2023). Cephalopod-omics: emerging fields and technologies in cephalopod biology. *Integrative and Comparative Biology*, *63*(6), 1226-1239. https://doi.org/10.1093/icb/icad087

Songco-Casey, J.O., **Coffing, G.C.,** Piscopo, D.M., Pungor, J.R., Kern, A.D., Miller, A.C., and Niell, C.M. (2022). Cell types and molecular architecture of the Octopus bimaculoides visual system. Current Biology *32*, 5031-5044.e4. https://doi.org/10.1016/j.cub.2022.10.015.

Battey, C. J., **Coffing, G. C**., & Kern, A. D. (2021). Visualizing population structure with variational autoencoders. *G3 11*(1), jkaa036. https://doi.org/10.1093/g3journal/jkaa036

Dyck, P. K. V., Hockaden, N., Nelson, E. C., Koch, A. R., Hester, K. L., Pillai, N., **Coffing, G. C.,** Burns, A.R., Lafontant, P.J. (2020). Cauterization as a Simple Method for Regeneration Studies in the Zebrafish Heart. *Journal of Cardiovascular Development and Disease, 7*(4), 41. MDPI AG. http://dx.doi.org/10.3390/jcdd7040041

Nelson, H. M.*, **Coffing, G. C.*,** Chilson, S., Hester, K., Carrillo, C., Ostreicher, S., Tomamichel, W., Hanlon, S., Burns, A. R., & Lafontant, P. J. (2019). Structure, development, and functional morphology of the cement gland of the giant danio, Devario malabaricus. *Developmental dynamics: an official publication of the American Association of Anatomists, 248*(11), 1155–1174. https://doi.org/10.1002/dvdy.88
* denotes co-first author

ACKNOWLEDGMENTS

I feel incredibly fortunate to have found a welcoming and supportive community in Oregon. I never imagined that I'd have the opportunity to live in such a beautiful place while studying one of the most extraordinary organisms on our planet. I am grateful to my supervisor, Andy Kern, for his mentorship, support, and constant enthusiasm. I admire his dedication to cultivating a welcoming and encouraging lab environment. I am thankful to my committee members, Cris Niell, Adam Miller, Mara Lawniczak, and Peter Ralph, for supporting me throughout my journey as a graduate student. I'm thankful to Cris and Adam for introducing me to cephalopod biology and letting me turn my rotation project into a Ph.D. project.

I want to sincerely thank all my lab members and colleagues—past and present—for the academic guidance and emotional support throughout my time as a graduate student. I'm grateful to the Octo crew for their support, patience, and enthusiastic collaboration early in my Ph.D. and for letting me join them at cephalopod conferences. I'm thankful to staff scientists Silas Tittes and Scott Small for their guidance and encouragement later in my Ph.D. journey. Outside of the lab, I'm grateful to Gilia Patterson for being a great friend and sharing so many memorable outdoor adventures with me.

I will forever be fortunate to have an incredible cohort including Aidan Short, Caitlin Smith, Hope Healey, and Victoria Caudill. I can't imagine graduate school without your unwavering support, positivity, and friendship. I'm so proud of you all and I can't wait to see what the future holds.

I am grateful for the support and guidance from my community in the Institute of Ecology and Evolution and the Department of Biology. The GrEBES student group has been a wonderful community of compassionate IE2 graduate students. I'm also grateful for the

leadership experience I gained through my involvement with Women in Graduate Sciences, and I'm continually inspired by the meaningful impact this supportive community has in Eugene. I'm grateful to the IE2 staff that answered all of my questions about conference travel, including Arlene Crain, Sara Nash, Maria Heider, and Leah Frazier.

I'm also deeply grateful to my undergraduate research mentor, Professor Pascal Lafontant and my undergraduate advisor, Professor Gloria Townsend, for their guidance and encouragement during my early academic journey. Their mentorship played a pivotal role in shaping me into a young researcher and equipping me with invaluable skills. I will always admire their dedication to undergraduate education.

Finally, I want to thank my loved ones for supporting me in pursuing my dreams. I'm thankful to Michael for consistently bringing joy into my life. Most importantly, I will always be grateful to my parents, Erin and Steve, for instilling in me a deep appreciation for the value of education and hard work.

DEDICATION

To my Dad

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I: INTRODUCTION

Cephalopods—including octopus, squid, cuttlefish, and nautilus—are a diverse class of molluscan invertebrates that have long captured the attention of both scientists and the public. Their remarkable collection of morphological innovations, from unique body plans and complex nervous systems to camera-type eyes and unparalleled camouflage abilities, have made cephalopods enduring figures in folklore, popular culture, and natural history media. The lineage of today's cephalopods traces back to the Early Cambrian period, roughly 550 million years ago, during a time of rapid evolutionary diversification[1–3]. Early cephalopods had external shells, leaving behind a rich fossil record that shows repeated speciation and extinction events. Around 480 million years ago, the coleoid cephalopods (octopus, squid, and cuttlefish), diverged from the nautiloids[1], giving rise to the modern groups that now inhabit all parts of the ocean from shallow reefs to the deep sea, with over 700 species described. Cephalopods have become a foundational study system in multiple areas of biology including neuroscience, behavior, biomechanics, evolution, and ecology. Notably, the squid giant axon played a pivotal role in the discovery of action potentials[4], which helped set the stage for research focused on the cephalopod brain and beyond.

Despite the rich history of cephalopod biology research, the field of cephalopod genomics is relatively nascent. Recent advances and declining costs of sequencing has facilitated the explosion of cephalopod genomics in recent years revealing that cephalopods also harbor numerous evolutionary novelties within their genomes. The first genome assembly of a cephalopod, the California two-spot octopus (*O. bimaculoides*), identified that cephalopod genomes are highly repetitive and they contain gene family expansions thought to underline the evolution of their complex nervous systems[5]. Additionally, cephalopods genomes contain

exceptionally long non-coding regions compared to other invertebrates[6]. Unlike any other organism, cephalopods extensively recode their neuronal and cytoskeletal expressed proteins with A-to-I RNA editing, suggesting it is important for brain physiology[7,8]. As more cephalopod genomes are assembled, comparative approaches can be used to examine their evolutionary history. These approaches have found that the coleoid genomes have extensive macro (chromosomal) and micro (local) genome rearrangements, unlike what is seen in other animals[9,10]. Together, these genomic insights are reshaping our understanding of the molecular foundations underlying cephalopod complexity and evolution.

My dissertation builds upon and leverages the recent explosion in cephalopod genomic data to learn more about the history of cephalopods on the macro- and micro- evolutionary scales. I generated and analyzed multiple types of sequencing data from the California two spot octopus, (*Octopus bimaculoides*) in three different projects. This data has allowed me to ask specific questions about the octopus visual system, cephalopod sex determination, and genomic variation within octopus populations.

In Chapter II, I present our efforts to determine the transcriptional cell types of the octopus visual system. In a textbook example of convergent evolution, octopuses and vertebrates have evolved anatomically similar camera-type eyes despite over 500 million years of independent evolution. Camera-type eyes give octopuses high acuity vision that is used for sophisticated visually driven behaviors. The overall organization of the octopus brain, however, is dramatically different than that of vertebrates and other invertebrates. Little is known about the cell types and molecular determinants of the octopus visual system and its neural circuit organization. To begin to address this, we aimed to generate a single-cell molecular atlas of the octopus optic lobe, the primary visual processing area of the octopus brain. Although *O.*

bimaculoides, our primary study organism, was the first cephalopod to have its genome sequenced in 2015, this assembly was highly fragmented, and thus problematic for in-depth transcriptional and single-cell analysis. To overcome this limitation, I generated a high-quality, chromosome-level genome assembly and annotation using long-read sequencing. As a result, we captured 13% more of the single-cell data in our analysis and we revealed six major neuronal cell classes identified by neurotransmitter expression as well as non-neuronal and immature neuron cell populations. This computational work guided RNA fluorescence in situ hybridization experiments that determined anatomical and spatial organization of cells within the optic lobe. This work was published in *Current Biology* and is co-authored by Dr. Jeremea O. Songco-Casey, Dr. Denise M. Piscopo, Dr. Judit R. Pungor, Dr. Andrew D. Kern, Dr. Adam C. Miller, and Dr. Cristopher M. Niell[11].

As I examined the chromosome-level octopus genome for Chapter II, I discovered a chromosome with half the coverage of sequencing reads compared to the rest of the genome. In Chapter III, I reveal that this half-coverage chromosome is a sex chromosome, the first evidence of genetic sex determination identified in a cephalopod. Using sequencing data from female and male individuals, I showed that the sex chromosome is hemizygous in females and diploid in males, meaning that *O. bimaculoides* has a ZZ/ZO karyotype. In other words, the male octopus has two copies of the Z chromosome and female octopus has only one. By examining the repeat landscape across chromosomes, I found that the Z has a significantly higher LINE element density compared to the rest of the genome. Curious whether this repeat pattern might be a signature of sex chromosomes in cephalopods, I looked at repeat densities in chromosome-level assemblies of other cephalopod species. I found that in each species I examined, there was a single chromosome with an elevated LINE density. By conducting a synteny analysis, I found

that the chromosomes with elevated LINE densities are orthologous across cephalopods, leading our team to call them putative sex chromosomes. I analyzed publically available coverage data in at least one species from each cephalopod clade (i.e. octopus, squid, cuttlefish, and nautilus) and found that the putative sex chromosome is half coverage for females (or presumably female) in each species. Altogether, my evidence suggests that the Z chromosome is shared across all extant cephalopods, placing its origin to be at least ~480 million years ago. This places the cephalopod sex determination system among the oldest known sex determination systems. This chapter has been published in *Current Biology* and is co-authored by Dr. Silas Tittes, Dr. Scott T. Small, Dr. Jeremea O. Songco-Casey, Dr. Denise M. Piscopo, Dr. Judit R. Pungor, Dr. Adam C. Miller, Dr. Cristopher M. Niell, and Dr. Andrew D. Kern[12].

The California two-spot octopus (*O. bimaculoides*) is a focal study species in cephalopod research. Interestingly, *O. bimaculoides* it was not recognized as a separate species from another two-spot octopus species (*O. bimaculatus*) until 1949 due to their similar appearances and overlapping geographic ranges[13]. In Chapter IV, I changed my focus of evolutionary processes from the macro to the micro scale in a population genetics study. I asked whether diversity, divergence, and demographic history of the *O. bimaculoides* and *O. bimaculatus,* a pair of sister, sympatric species, is consistent with their life history and reproductive strategies. *O. bimaculoides* and *O. bimaculatus* have overlapping geographic ranges off the coast of southern California through northern Mexico. Despite their similar morphological appearances, their reproductive strategies are quite distinct. *O. bimaculoides* lay relatively few eggs that develop directly into benthic hatchlings where *O. bimaculatus* produce a comparatively high number of eggs that have a planktonic paralarvae stage where they disperse far distances from their hatching site. Our team conducted whole genome sequencing on a set of individuals from each

species. With this data, I called variant sites and calculated population genetics statistics including nucleotide diversity (π), Tajima's D, d_{XY}, and F_{ST}. These results show that nucleotide diversity is lower in *O. bimaculoides* than *O. bimaculatus* and that d_{XY} estimates show few fixed differences between the species. I compared demographic history between the species to show that *O. bimaculatus* has historically had a larger effective population size than *O. bimaculoides*. Finally, I found both species have similar recombination rates, however the overall window mean estimates are significantly higher in *O. bimaculoides.* This chapter is co-authored with Dr. Silas Tittes, Dr. Scott Small, and Dr. Andrew Kern.

Together, my dissertation explores the evolutionary past of an emerging model octopus species. Chapter II introduces a chromosome-level genome assembly and annotation, which was essential for generating one of the first cephalopod brain transcriptome atlases. Chapter III discusses my novel discovery of the cephalopod sex determination system, which had been a long-standing mystery. Chapter IV highlights the first whole genome resequencing dataset in a cephalopod to describe the history of an octopus species important to both the research and fishery communities.

References

1. Huang, Z., Huang, W., Liu, X., Han, Z., Liu, G., Boamah, G.A., Wang, Y., Yu, F., Gan, Y., Xiao, Q., et al. (2022). Genomic insights into the adaptation and evolution of the nautilus, an ancient but evolving "living fossil." Molecular Ecology Resources *22*, 15–27. https://doi.org/10.1111/1755-0998.13439.

2. Ponder, W. ed. (2008). Phylogeny and Evolution of the Mollusca (University of California Press) https://doi.org/10.1525/california/9780520250925.001.0001.

3. Chen, Z., Baeza, J.A., Chen, C., Gonzalez, M.T., González, V.L., Greve, C., Kocot, K.M., Arbizu, P.M., Moles, J., Schell, T., et al. (2025). A genome-based phylogeny for Mollusca is concordant with fossils and morphology. Science *387*, 1001–1007. https://doi.org/10.1126/science.ads0215.

4. Hodgkin, A.L., and Huxley, A.F. (1939). Action Potentials Recorded from Inside a Nerve Fibre. Nature *144*, 710–711. https://doi.org/10.1038/144710a0.

5. Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., and Rokhsar, D.S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature *524*, 220–224. https://doi.org/10.1038/nature14668.

6. McCoy, M.J., and Fire, A.Z. (2020). Intron and gene size expansion during nervous system evolution. BMC Genomics *21*, 360. https://doi.org/10.1186/s12864-020-6760-4.

7. Alon, S., Garrett, S.C., Levanon, E.Y., Olson, S., Graveley, B.R., Rosenthal, J.J.C., and Eisenberg, E. (2015). The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. eLife *4*, e05198. https://doi.org/10.7554/eLife.05198.

8. Voss, G., and Rosenthal, J.J.C. (2023). High-level RNA editing diversifies the coleoid cephalopod brain proteome. Briefings in Functional Genomics *22*, 525–532. https://doi.org/10.1093/bfgp/elad034.

9. Albertin, C.B., Medina-Ruiz, S., Mitros, T., Schmidbaur, H., Sanchez, G., Wang, Z.Y., Grimwood, J., Rosenthal, J.J.C., Ragsdale, C.W., Simakov, O., et al. (2022). Genome and transcriptome mechanisms driving cephalopod evolution. Nat Commun *13*, 2427. https://doi.org/10.1038/s41467-022-29748-w.

10. Schmidbaur, H., Kawaguchi, A., Clarence, T., Fu, X., Hoang, O.P., Zimmermann, B., Ritschard, E.A., Weissenbacher, A., Foster, J.S., Nyholm, S.V., et al. (2022). Emergence of novel cephalopod gene regulation and expression through large-scale genome reorganization. Nat Commun *13*, 2172. https://doi.org/10.1038/s41467-022-29694-7.

11. Songco-Casey, J.O., Coffing, G.C., Piscopo, D.M., Pungor, J.R., Kern, A.D., Miller, A.C., and Niell, C.M. (2022). Cell types and molecular architecture of the Octopus bimaculoides visual system. Curr Biol *32*, 5031-5044.e4. https://doi.org/10.1016/j.cub.2022.10.015.

12. Coffing, G.C., Tittes, S., Small, S.T., Songco-Casey, J.O., Piscopo, D.M., Pungor, J.R., Miller, A.C., Niell, C.M., and Kern, A.D. (2025). Cephalopod sex determination and its ancient evolutionary origin. Curr Biol *35* 931-939.e4. https://doi.org/10.1016/j.cub.2025.01.005.

13. Pickford, G., and McConnaughey, B. (1949). The Octopus bimaculatus problem: a study in sibling species. Bulletin of the Bingham Oceanographic Collection *12*.

II: CELL TYPES AND MOLECULAR ARCHITECTURE OF THE *OCTOPUS BIMACULOIDES* VISUAL SYSTEM

This chapter was published in the journal *Current Biology* in 2022. Co-authors include Dr. Jeremea O. Songco-Casey, Dr. Denise M. Piscopo, Dr. Judit R. Pungor, Dr. Andrew D. Kern, Dr. Adam C. Miller, and Dr. Cristopher M. Niell. C.M.N. and A.C.M. conceived and oversaw the project. J.O.SC., D.M.P., and J.R.P. collected sequencing data. G.C.C. and A.D.K. performed genome assembly and annotation, J.R.P. performed cross-species gene identification, J.O.SC., A.C.M., and C.M.N. performed scRNA-seq analysis. J.O.SC. and D.M.P. performed FISH experiments. All authors contributed to writing and editing of the manuscript.

The citation for this publication is as follows:

Songco-Casey, J. O., **Coffing, G. C**., Piscopo, D. M., Pungor, J. R., Kern, A. D., Miller, A. C., & Niell, C. M. (2022). Cell types and molecular architecture of the Octopus bimaculoides visual system. *Current Biology*, 32(23), 5031-5044. https://doi.org/10.1016/j.cub.2022.10.015

Introduction

Cephalopods represent a unique branch of the animal kingdom for studying vision. Coleoid cephalopods (octopuses, squids, and cuttlefish) have the largest brain among invertebrates[1–3], much of which is comprised of areas dedicated to visual processing: the optic lobes[4,5]. Their visual system facilitates a range of behaviors such as navigation, prey capture, and complex camouflage[1,6–8]. However, the neural basis of central visual processing in cephalopods is largely unknown. Despite diverging over 500 million years ago, octopuses independently evolved camera-type eyes similar to those of vertebrates[9]. However, the neural organization of

structures that process visual information is dramatically different in the two lineages. In contrast to the vertebrate retina, which is an intricate neural circuit with a diversity of cell types, the octopus retina consists only of photoreceptors and supporting cells[10,11]. The photoreceptors send axons into the central brain, targeting the optic lobes that lie directly behind the eyes[10] (Figure 2.8A). The optic lobes, which make up approximately two-thirds of the octopus' central brain[12,13], are where the bulk of visual processing is hypothesized to occur[12–14]. Optic lobe outputs project to deeper brain regions, including those involved in learning/memory and motor behavior[12,15–19].

Histological studies by Young[5] have provided a description of the optic lobe's anatomical organization and neuronal morphologies (Figure 2.1B-D), which we briefly summarize here. The outermost region of the optic lobe is a cell body layer, termed the outer granular layer (OGL). The OGL contains cells traditionally referred to as amacrine cells based on their morphology[18,20], which have multipolar processes that ramify within the plexiform layer (PL) below the OGL. The PL is a dense neuropil and is the primary termination site of photoreceptor axons, optic lobe neuronal processes, and projections from deeper brain regions[12,20]. Below the PL is another cell body layer, the inner granular layer (IGL), which has a varied population including (1) neurons with bipolar morphology, (2) neurons that send centrifugal axons back to the retina, and (3) neurons of amacrine morphology with processes in the plexiform layer[12,18]. Finally, a deeper structure termed the medulla comprises the bulk of the optic lobe. The cells within the medulla are organized in a branching tree-like fashion which, in cross-section, appear as islands of cell bodies surrounded by neuropil[21]. The superficial region of the medulla is organized into columns

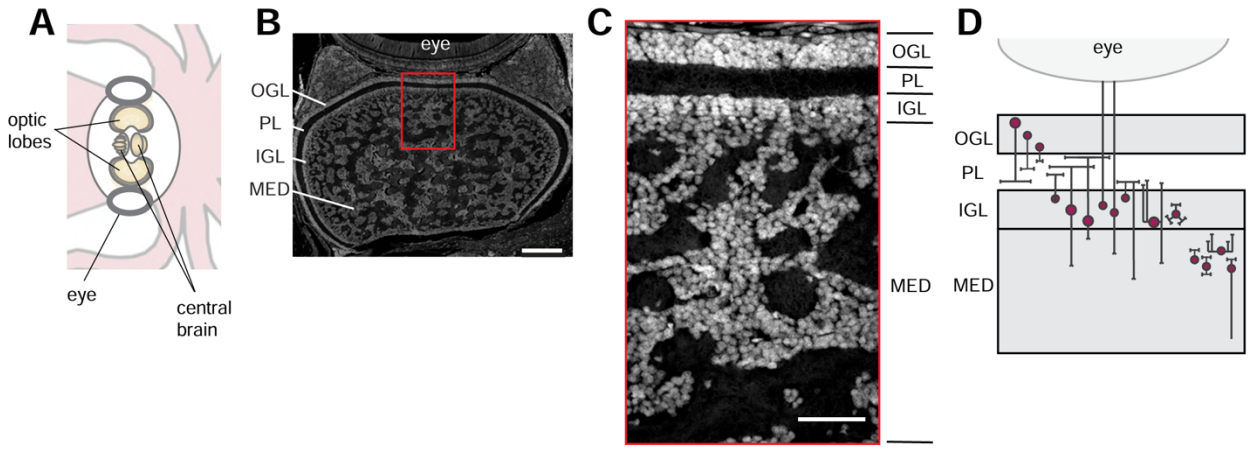

Figure 2.1: Laminar organization of *O.bimaculoides* optic lobes
(A) Schematic showing the octopus central nervous system, which includes optic lobes behind each eye and the central brain in between.
(B) Overview of one optic lobe and eye. Nuclei are stained with DAPI. Scale bar indicates 200 um. Red box denotes the region shown in C. OGL, outer granular layer; PL, plexiform layer; IGL, inner granular layer; MED, medulla.
(C) Laminar organization of the optic lobe demonstrated by nuclear staining of a cross-section. Scale bar indicates 50 um.
(D) Schematic of the anatomical organization of the visual system, in terms of neuronal morphologies, adapted from [5].

and is referred to as the outer radial columnar zone, whereas the deeper region of the medulla includes processes that extend tangentially and is termed the central tangential zone[5]. Given the anatomical organization, it has been hypothesized that the outer layers of the optic lobe (OGL, PL, IGL) may perform similar functions to the vertebrate retina, leading it to be termed the "deep retina"[22], while the medulla may engage in higher order processing analogous to central visual areas in other species[12].

While the anatomy of these cell classes suggest an organizational foundation, fully understanding the neural circuitry necessitates knowledge of the molecular identities of these cell types, including both functional determinants (e.g. neurotransmitter and receptor repertoires) and developmental determinants (e.g. transcription factors and adhesion molecules). Recent molecular taxonomies in other species have provided new insight into the circuit organization of

a number of brain regions including the fly visual system[23], the mouse and primate retina[24,25] and the mouse visual cortex[26]. We therefore sought to create a systematic molecular characterization of the octopus visual system by combining single-cell RNA sequencing (scRNA-seq), to determine transcriptional cell types, with multiplexed RNA fluorescence in situ hybridization (FISH), to determine the location of these cell types within the optic lobe.

Results

ScRNA-seq of the octopus optic lobe

We performed scRNA-seq in juvenile (1.5 months of age) Octopus bimaculoides. Despite their continued growth throughout their lifetime (1-2 years), the overall body organization, behavior, and visual system of *O. bimaculoides* are mature by this age[27,28], allowing us to identify neural circuitry involved in a fully functioning, yet still growing, visual system. We performed Chromium 10x sequencing of cells from the optic lobes of two animals, processed separately as biological replicates, and aligned reads to an updated genome assembly and gene annotation using CellRanger (see Methods, Figure S2.1 and Table S2.1).

A robust scRNA-seq analysis requires a contiguous genome with accurate, full-length gene annotations. The first genome assembly of *O. bimaculoides* (Octopus_bimaculoides_v2_0) is broken into over 150,000 scaffolds, with many split or truncated gene annotations. Alignment of our single cell reads to this genome therefore resulted in low mapping. To resolve this, we used single molecule high fidelity (HiFi) sequencing to create a new contig level genome assembly and combined new Iso-seq reads with existing bulk RNA-seq data to generate an improved genome annotation (see Methods, Figure S2.1 and Table S2.1) containing 5,437 contigs and 18,896 gene annotations. The new assembly is more contiguous and helped lengthen

the 3' ends of many annotated genes, capturing more reads (Figure S2.1) and thus achieving higher resolution.

Using standard filtering, normalization, integration and clustering in Seurat[29,30], we identified a total of 28,855 cells across two biological replicates. Our analysis resulted in a total of 41 clusters, where each sample contributed cells to all of the identified clusters in similar proportions, supporting the reproducibility of this approach (Figure S2.2).

We first sought to broadly characterize the identified clusters in terms of neuronal and non-neuronal populations (Figure S2.3). We used a homologous sequence identifier, OrthoFinder[31], to assign gene-family and orthology relationships using genes from Drosophila, vertebrates, and other cephalopod species (see Methods, Figure S2.3A for example gene trees). Genes that were not assigned to orthology groups by OrthoFinder were manually annotated using NCBI BLAST[32] if possible. Throughout, we name the octopus genes according to their assigned identity: e.g. synaptotagmin (syt), as summarized in Table S2.2, although we note that these assignments may be improved as our understanding of gene homology in cephalopods advances.

We expected a large proportion of the cells to have high expression of genes related to mature neurons, as well as potentially developing neurons, as the octopus brain continues to grow and add neurons throughout its lifetime[33]. To identify neurons within our scRNA-seq data, we first looked at the expression of genes related to synaptic vesicle release. We found that most clusters (33/41) expressed the *O. bimaculoides* SNARE genes, while only seven were likely to be non-neuronal, based on the absence of expression of these markers (Figure S2.3). The non-neuronal clusters represented ~8% of the cells and had relatively high expression of genes falling within gene families with functions consistent with proliferation, blood, endothelium, and glia (Figure S2.3). We used FISH to localize expression of non-neuronal genes and found several to

be primarily expressed outside of the optic lobe (see Figure S2.3 for further characterization). We therefore removed these clusters from subsequent analyses that aimed to delineate neuronal cell types, which included a total of 26,092 cells across 33 clusters.

To further explore the potential neural cell type classes within the scRNA-seq clusters, we first examined the expression patterns of markers for neurotransmitter usage (Figure 2.2). Previous work identified dopamine, glutamate, and acetylcholine as the primary neurotransmitters used in the cephalopod optic lobe[34], so we looked for standard markers of these neurotransmitter types based on their biosynthetic and vesicular packaging pathways (dopamine transporter (*dat*), vesicular glutamate transporter (*slc17a7*, or *vglut*), vesicular acetylcholine transporter (*slc18a3*, or *vacht*), and choline acetyltransferase (chat)) (Figures 2.2B-C, S2.4). In addition, we identified a gene in the solute carrier gene superfamily (*slc6a15/18*), which acts to support glutamate synthesis via transport of glutamate precursors[35] and was closely co-expressed with vglut in both scRNA-seq data (Figure 2.2C) and FISH (Figure S2.4). We used *slc6a15/18* as an additional marker of glutamatergic neurons along with vglut, since, despite strong FISH signal, we found a relatively low number of scRNA-seq reads aligned to vglut, which may be due to an incomplete gene model for this gene.

Together, scRNA-seq expression of neurotransmitter markers delineated the majority of putative neurons (22/33 clusters) into four broad classes defined by either unique or combinatorial expression of these genes (Figure 2.2B-C). Each of these broad categories consisted of a number of unique clusters (Figure 2.2A, C), suggesting further cell type heterogeneity within. In addition, two smaller neuronal clusters were identified, one of which did

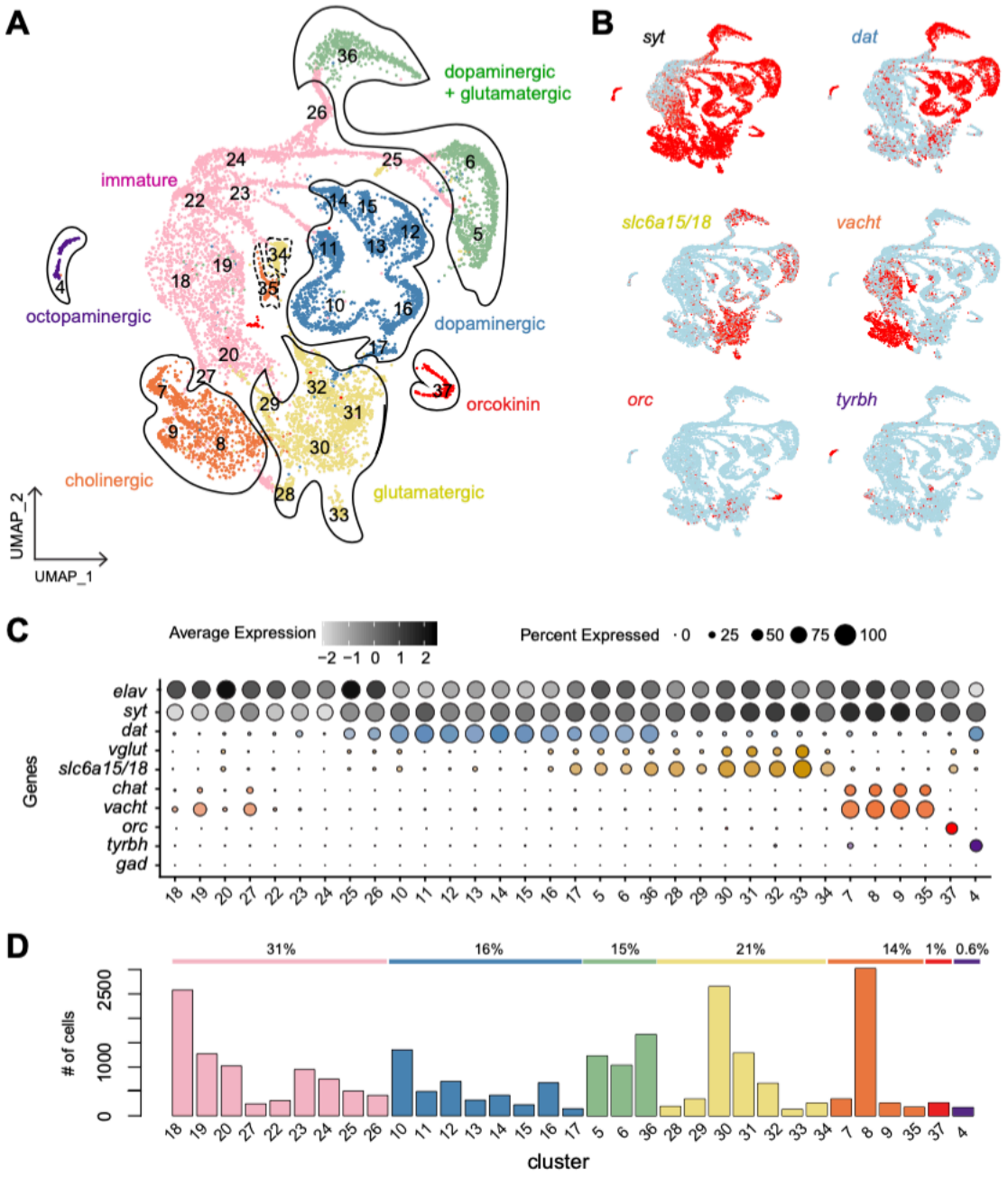

Figure 2.2: scRNA-seq reveals six major neuronal classes
(A) UMAP of putative neuronal clusters. A total of 33 clusters are color-coded based on major cell classes.
(B) Feature plots showing expression patterns of marker genes for neurons (syt) and neurotransmitter phenotypes (dopaminergic cells (dat), glutamatergic cells (slc6a15/18), cholinergic cells (vacht), orcokinin cells (orc), and octopaminergic cells (tyrbh).
(C) Dot plot of markers delineating molecular cell classes.

(D) Bar graph indicating total number of cells in each cluster as colored in (A) as well as the relative proportion of each cell class across the entire population of neurons in the optic lobe.
See also Figures S2.1-S2.4, Tables S2.1-S2.2

not express markers for any neurotransmitters, but did express the neuropeptide orcokinin (*orc*) (Figure 2.2B) and another that expressed a combination of dat and a marker for octopamine synthesis, tyramine beta-hydroxylase (*tyrbh*), previously identified in octopus optic lobe neurons[36]. We did not find a significant population of GABAergic neurons (based on expression of glutamate decarboxylase (gad), Figure 2.2C), consistent with previous findings of the minimal role of GABA in the optic lobe[37].

Finally, we found a large population of cells (9 clusters) that appear to be immature neurons, based on higher expression levels of early neural specification genes (i.e. embryonic lethal abnormal vision (*elav*) and CUG triplet repeat RNA binding protein (*celf*)), lower levels of genes involved in synaptic transmission, and no expression of any neurotransmitter markers (Figures 3.2C, S2.3). As described below, this class of cells expressed a diversity of evolutionarily conserved developmental genes and distinct subgroups had expression profiles suggestive of a relationship to distinct mature cell clusters.

Taken together, these findings support the idea that the scRNA-seq data captured expression profiles of unique classes of mature and immature neurons in the optic lobe. We used these data to delineate molecular cell types and assign them to an anatomical organization within the octopus optic lobe.

<u>A molecular and spatial taxonomy of mature neural cell types</u>

Our scRNA-seq data show that neurotransmitter usage divides the majority of octopus optic lobe neurons into four large classes – dopaminergic, dopaminergic+glutamatergic, glutamatergic, and cholinergic neurons – along with two smaller classes that utilize orcokinin or octopamine. To examine the localization of the four major neurotransmitter cell clases within the optic lobe, we performed FISH for dat, *slc6a15/18*, and vacht (Figure 2.3). Each of these neurotransmitter markers showed a distinct pattern of expression within the cell body layers of the OGL, IGL, and medulla (Figure 2.3B). Dopaminergic (*dat*+) cells were found predominantly in the OGL, with sparser expression in the IGL and medulla. Glutamatergic cells (*slc6a15/18*+, and see vglut in Figure S2.4) were found across the optic lobe, including in some dat+ cells in the OGL and IGL. Cholinergic cells (*vacht*+, and see chat in Figure S2.4) were restricted to the IGL and the medulla. As suggested by the scRNA-seq data (Figure 2.3A), glutamatergic and cholinergic neurons are segregated in their spatial expression patterns – both are expressed in non-overlapping cells in the IGL and medulla, with *slc6a15/18*+ expressed more strongly in the tangential region of the medulla and vacht+ expressed more strongly in the radial columnar region (Figure 2.3B). These mappings confirm that the scRNA-seq data identified distinct populations of neurons that correlate with distinct anatomical spatial expression patterns.

Within each of the large neurotransmitter-defined populations, further transcriptional heterogeneity was present, suggesting that the clusters within these populations may correspond to neuronal subtypes (Figure 2.2). We sought to determine if such cell types, as identified by their unique gene expression profiles, would occupy distinct anatomical locations within the optic lobe.

Dopaminergic neurons

In the scRNA-seq data, dopaminergic neurons spanned seven clusters (Figure 2.2), and dat+-only neurons were predominantly localized to the OGL (Figure 2.3B). We examined gene expression across dat+ clusters and found two subgroups defined by the complementary expression of either the homeobox transcription factor *six4/5* (clusters 12-17) or the neuropeptide fmrf (clusters 10-11) (Figure 2.4A). Across model species, the six family of genes are key regulators in head development[38], eye specification, and retinal determination[39–41]. fmrf neuropeptides are known to regulate a variety of functions in mollusks[42], including reproduction[43] and chromatophore control[44,45]. FISH revealed that within the dat+ cells in the OGL, the *six4/5*+ population represented a broad sub-layer of neurons in the middle of the OGL, while fmrf expression corresponded to a sub-layer of neurons deeper in the OGL, along the border of the plexiform layer. Thus, the *dat*+-only cells contain two subtypes that are differentiated by expression of a homeobox transcription factor and a neuropeptide respectively. Notably, these are mainly localized within the OGL and hence likely represent a subset of the amacrine neurons of Young's anatomical classifications[5].

Expression heterogeneity of additional genes suggested these two dopaminergic cell groups can be further subdivided. The clusters within the *six4/5*+ group differentially express several genes encoding neuropeptides (*pxfv, lxgkr,* and *flri*) which are largely non-overlapping, though FISH reveals a low level of co-expression (Figure 2.4A, see Figure 2.7D). This set of neuropeptides is particularly interesting as they were manually assigned identities based on repetitive protein sequences, but there is little information regarding their function in other organisms, let alone cephalopods. Further, a subset of cells within the fmrf+ group expresses the adhesion molecular dscam, which has been shown to mediate cell-type specific self-avoidance among dopaminergic amacrine cells[46] and sublayer-specific connectivity in amacrine and bipolar

cells[47] in the vertebrate retina. dscam+ cells form a discrete narrow band along the deep border of the OGL and the plexiform layer (Figure 2.4A), suggesting dscam could play a role in assigning their sublayer specificity. Together, these data demonstrate further heterogeneity within dopaminergic cell types in the OGL, identifying laminar organization from the superficial to deep layers.

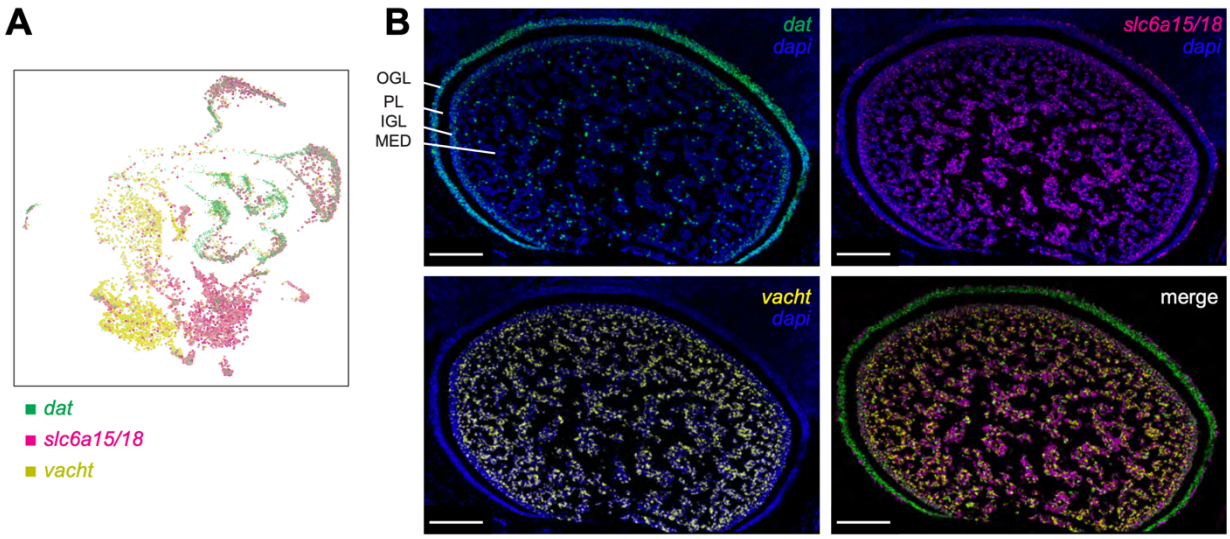

Figure 2.3: Neurotransmitter usage divides the majority of cells into four large populations
(A) UMAP of overlay of dat, slc6a15/18, and vacht expression.
(B) FISH of dat, slc6a15/18, vacht, with nuclei stained in DAPI, and merged FISH of the three neurotransmitter-related genes. Scale bars represent 100um. Here and below, merged images are shown without DAPI to emphasize relative expression patterns.
See also Figure S2.4, Table S2.2

Dopaminergic + glutamatergic neurons

Based on the co-expression of dat and slc6a15/18 in both the scRNA-seq and FISH data, we sought to further delineate this cell class. The scRNA-seq clusters with overlapping dat and slc6a15/18 expression suggested that two prominent groups, clusters 5/6 and 37, might correspond to subtypes. Furthermore, the expression of dat and *slc6a15/18* significantly overlapped in both the OGL and IGL, suggesting that these two locations might correspond to

the two groups (Figure 2.3A-B). The currently uncharacterized gene *obimac0010569* (see Methods S2.1 for further information) was uniquely expressed in cluster 37, and FISH revealed that these cells were located in the OGL (Figure 2.4B). This represents an additional population of OGL neurons, beyond the dopaminergic neurons discussed above. On the other hand, a nicotinic acetylcholine receptor (*nachr*) was expressed in clusters 5/6, and these were localized to a broad band in the IGL (Figure 2.4B). Together, these data demonstrate that the scRNA-seq dopamine+glutamatergic clusters consist of two distinct subtypes, one in the OGL (*obimac0010569*+) and one in the IGL (*nachr*+).

Glutamatergic neurons

We next focused on the subtypes of putative glutamatergic neurons, which include several smaller clusters (33-34, 28) in addition to a set of larger clusters (29-32) (Figure 2.2A). Examining genes in the smaller clusters revealed that these contain further subtypes. The first cluster (33) was defined by hedgehog (hh), a signaling molecule involved in axon guidance and patterning in the nervous system across many species[48]. FISH showed that hh+ cells are mainly restricted to a narrow band of neurons in the most superficial OGL, identifying yet another subtype within the OGL (Figure 2.4C). A second cluster (34) is marked by the voltage gated calcium channel gamma subunit 5/7 (*cacng*), which FISH demonstrated corresponds to a narrow band of neurons at the border of the IGL and medulla (Figure 2.4C). Finally, a third cluster (28) specifically expressed a member of the leucine-rich repeat family of cell adhesion molecules (*lrrc15*), which is involved in cell-type specific synaptic connectivity in fly and mammalian nervous systems[49]. This group was localized to the deeper region of the medulla (Figure 2.4C). Within the larger set of glutamatergic clusters (29-32), we found that a subset of neurons express

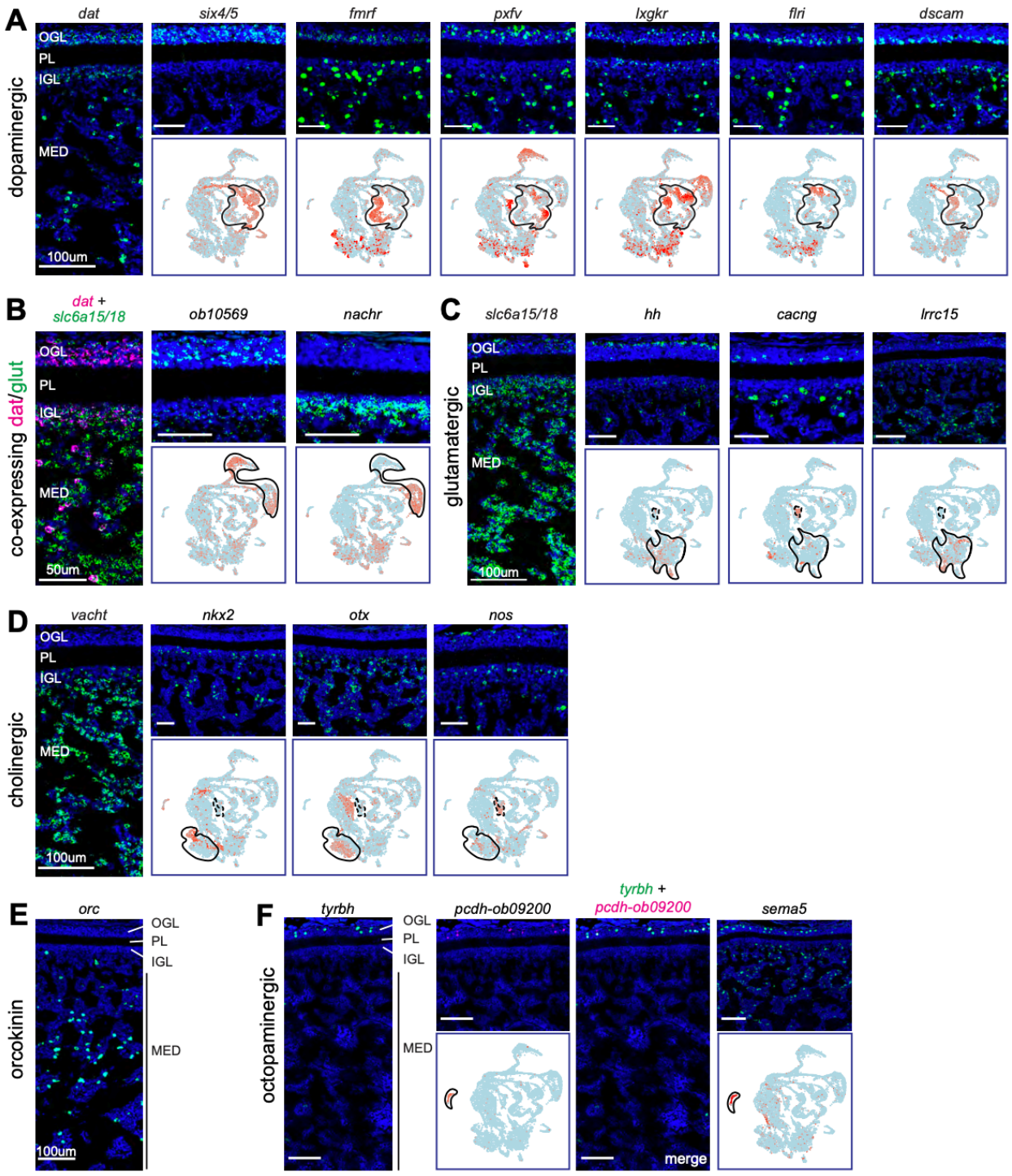

Figure 2.4: Anatomical organization of major cell classes and subtypes based on scRNA-seq and FISH

(A) Dopaminergic neuron organization. dat+ cells are divided into two major subtypes based on the additional differential expression of either six4/5+ or fmrf+, depicted through FISH and scRNA-seq feature plots. Corresponding cell classes are outlined in black on the feature plots, nuclei are stained with DAPI (blue), and scale bars represent 50um.

(B) Neuron organization for scRNA-seq clusters that co-express dat and slc6a15/18.

(C) Glutamatergic subtype neuron organization.
(D) Cholinergic subtype neuron organization.
(E) Orcokinin neuron organization.
(F) Octopaminergic neuron organization. From left to right: tyrbh, pcdh-obimac0009200, double FISH of tyrbh and pcdh-obimac0009200, and a single FISH of sema5.
See also Table S2.2, Supplemental Text

vesicular amine transporter 1 (*slc18a1,* or *vat1*), which FISH demonstrated to also be localized to cell bodies in the deeper region of the medulla (Figure 2.7F). Thus, the glutamatergic neurons primarily constitute a large population of cells within the medulla, along with two subtypes with highly specific sublayer localization within the outer OGL and inner IGL.

Cholinergic neurons

The last neurotransmitter class of cells in the octopus optic lobe is the putative cholinergic neurons. scRNA-seq data revealed a large population of cells (clusters 7-9) and a much smaller population (cluster 35) that both express vacht, a marker for cholinergic transmission. FISH for vacht showed that cholinergic neurons are located throughout the medulla, with more restricted expression in IGL (Figures 3.3B, 2.4D). Moreover, the scRNA-seq data revealed that clusters that comprise this larger population can be delineated based on the expression of family members of two homeobox transcription factors, *nkx2* and *otx*. FISH for these two markers demonstrated that the *nkx2*+ population is located in the IGL and superficial region of the medulla, while the otx+ population is not expressed in the IGL and, instead, is found throughout the medulla (Figure 2.4D). In exploring the scRNA-seq data, we found that the *nkx2*+ cluster also expresses a protocadherin (*obimac0026462*), and FISH confirms that *obimac0026462*+ cells are expressed both in the IGL and medulla (Figure 2.7E).

We observed that neurons in the smaller cholinergic cluster (35) selectively express nitric oxide synthase (*nos*), suggesting that they represent a distinct cell type that uses this neuropeptide as a signaling molecule in addition to acetylcholine. FISH for nos revealed these neurons form a narrow sub-layer within the superficial to central IGL (Figure 2.4D). scRNA-seq data show that nos+ cells and the glutamatergic *cacng*+ cells, both in the IGL, also express an unidentified gene *obimac0022194* (see Methods S2.1), suggesting that these clusters may have some shared function based on expression of this unidentified gene. Thus, the cholinergic neurons constitute a large population of the IGL and medullar cells, with distinct anatomical positions defined by a small sub-layer of *nkx2*+ cells in the IGL and superficial medulla, *otx*+ cells throughout the medulla, and nos+ cells mainly in the superficial IGL.

Orcokinin and Octopaminergic neurons

We sought to determine the identity of the two remaining mature neuronal clusters (37 and 4), which were defined by highly specific expression of *orc* and *tyrbh*, respectively (Figure 2.2A-C). Cluster 37 did not express any of the neurotransmitter markers we assessed, but was demarcated by the expression of the neuropeptide orc, and FISH revealed that *orc*+ cells are a sparse, scattered population throughout the deeper region of the medulla (Figure 2.4E). The final cluster of mature neurons, cluster 4, selectively expressed a number of genes, including tyrbh, the synthetic enzyme for octopamine, generally considered to be the invertebrate analog of norepinephrine[50], which plays a role in arousal and other aspects of behavioral state across species. FISH showed a discrete population of octopaminergic tyrbh+ neurons in the OGL (Figure 2.4F). Among the other genes that were unique to this cluster is a protocadherin family member (*obimac0009200*), which co-expresses with tyrbh exclusively in the OGL (Figure 2.4F),

and semaphorin-5 (*sema5*), which is highly expressed in the same region of the OGL, along with some expression throughout the medulla (Figure 2.4F). Thus, this cluster also expresses genes that serve as adhesion molecules (*protocadherin obimac0009200*) and axon guidance cues (*sema5*) in the visual system of other species[51,52].

Immature Neurons

Finally, we examined the putative immature neuronal clusters, representing 31% of all neurons, which were identified in the scRNA-seq data based on lower expression of syt and the absence of mature neurotransmitter markers. We observed that in the U-MAP these clusters were organized into several 'arms', suggesting that they may represent discrete populations of developing neurons associated with mature cell types (Figure 2.2A). One unidentified gene (*obimac0011980*) appears to encompass all nine immature clusters (Figure 2.5A-B, see Methods S2 for further characterization). The immature clusters can then be segregated into three subgroups that are demarcated by complementary expression of genes: a previously unidentified gene obimac0032150 which contains a tumor necrosis factor receptor domain (*tnfr*; clusters 18-20, 27), myoneurin (*mynn*; clusters 22-24), and big brain (*bib*; clusters 25-26) (Figure 2.5A-B), all of which are known to play a role in neural development in other species. Since these three genes segregate the population of immature neurons, we used FISH to identify the expression and examine their relationships to the location of the mature cell types described above (Figure 2.5C).

In the first sub-group, tnfr expression appeared in scRNA-seq clusters that are transcriptionally related to mature clusters of cholinergic and glutamatergic neurons (Figures 3.2A, 2.5B), which are largely found in the medulla. FISH data for tnfr revealed more extensive

expression in the medulla, compared to the other putative immature subtypes (Figure 2.5C). In the second sub-group, scRNA-seq data showed that mynn, a zinc finger protein family member associated with neuromuscular synapse formation in mice[53], is expressed in a number of clusters in the “arms" leading to mature cell types for dopaminergic neurons of the OGL (Figures 3.2A, 2.5B). Correspondingly, FISH data shows that the mynn+ cell types border the plexiform layer along both the OGL and IGL, together with some cells in the medulla (Figure 2.5C). In the third subgroup, bib, a known neurogenic molecule in Drosophila[54], was expressed in “arms” leading to the clusters that correspond to the two prominent clusters of dopaminergic+glutamatergic neurons (Figures 3.2A, 2.5B). We found *bib* was expressed most strongly, though not exclusively, in cells along the bottom borders of the IGL and OGL (Figure 2.5C). As mentioned above, all of these subgroup markers are also expressed throughout the medulla, suggesting they may represent the ongoing migration of immature neurons into the optic lobe. They also have increased expression in stratifications along the borders of the OGL and IGL, raising the possibility that the laminar borders may be an important locus for the incorporation of immature neurons.

Other known developmental genes (Figure 2.5A, D) were expressed in the immature clusters in the scRNA-seq data, and we investigated the expression patterns of three of these genes using FISH: *dschs, sox2,* and *dlx* (Figure 2.5E). We also identified two receptor-ligand pairs that have well established roles in patterning the nervous system in vertebrates and other invertebrates[51,55] - ephrin (*efn*)/Eph receptor (*epha*) and semaphorin-2 (*sema2*)/plexin (*plxna*) - all of which had complementary expression patterns to each other across the scRNA-seq data (Figure 2.6A-B). Graded expression of ephrins and semaphorins, and their respective receptors,

play important roles in establishing large-scale organization during development in other species, including topographic map formation[51,55]. Notably, the ligand-receptor pair efn and epha were

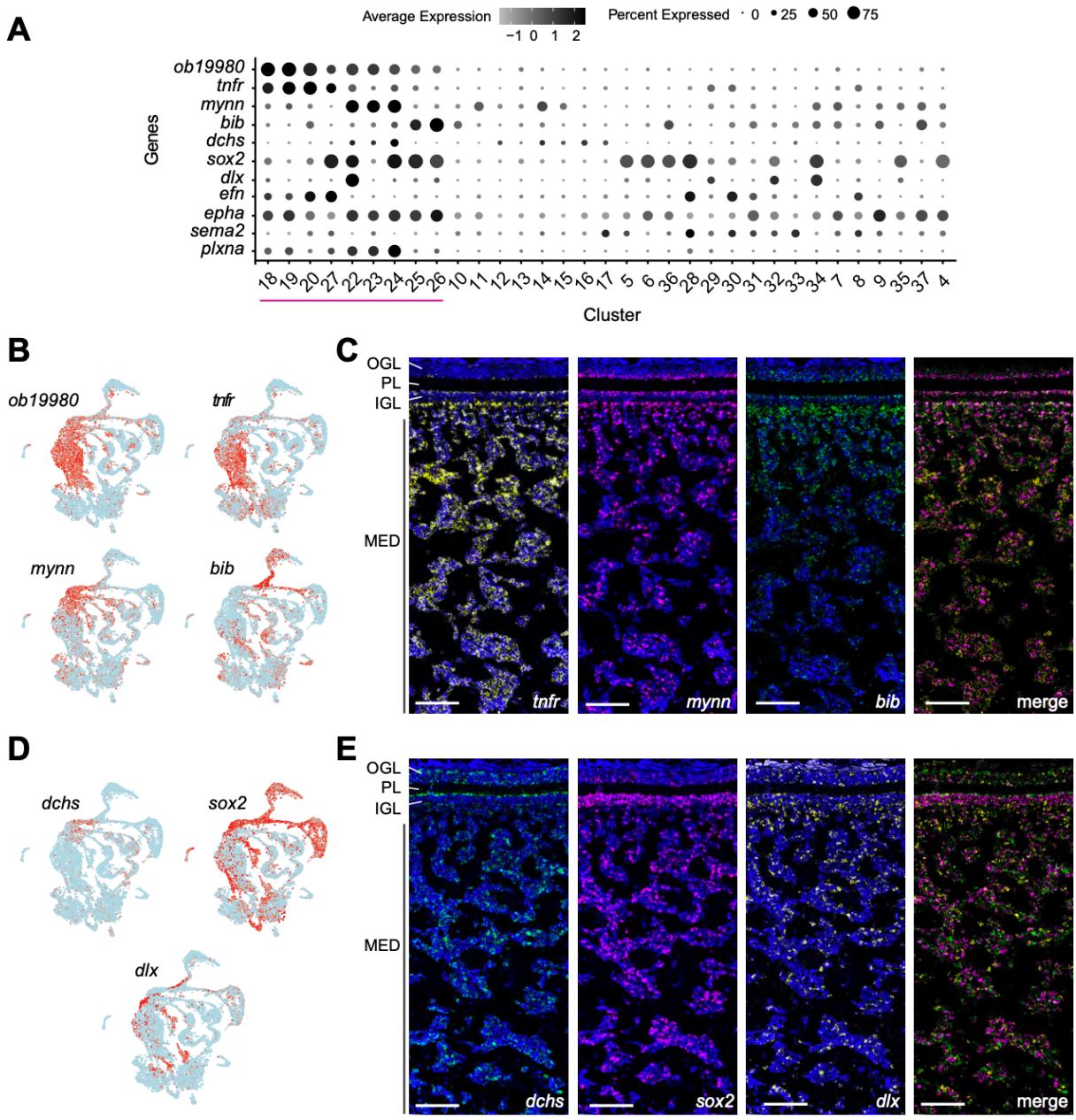

Figure 2.5: Gene expression and spatial organization of putative immature neurons
(A) Dot plot of genes expressed in immature neurons, with magenta line highlighting immature clusters.
(B) Feature plot of uncharacterized cephalopod-specific gene obimac0019980, which demarcates the putative immature neuronal clusters, as well as three genes that define distinct subgroups within the immature neurons: tnfr, mynn, and bib.
(C) FISH of the genes delineating the three subgroups shown in B, including a merged FISH. Throughout this figure, DAPI is shown in blue on individual FISHs, and scale bars indicate 100um.

(D) Feature plots of additional markers from development-related gene family trees demonstrating further cell type diversity.
(E) FISH showing anatomical organization of the genes shown in D, including a merged FISH.
See also Table S2.2, Supplemental Text

expressed in opposing gradients from superficial to deep in the optic lobe (Figure 2.6A). Likewise, sema2 was expressed in a gradient with strongest expression in the deep medulla, while its potential receptor plxna had stronger expression along the borders of the OGL and IGL (Figure 2.6B).

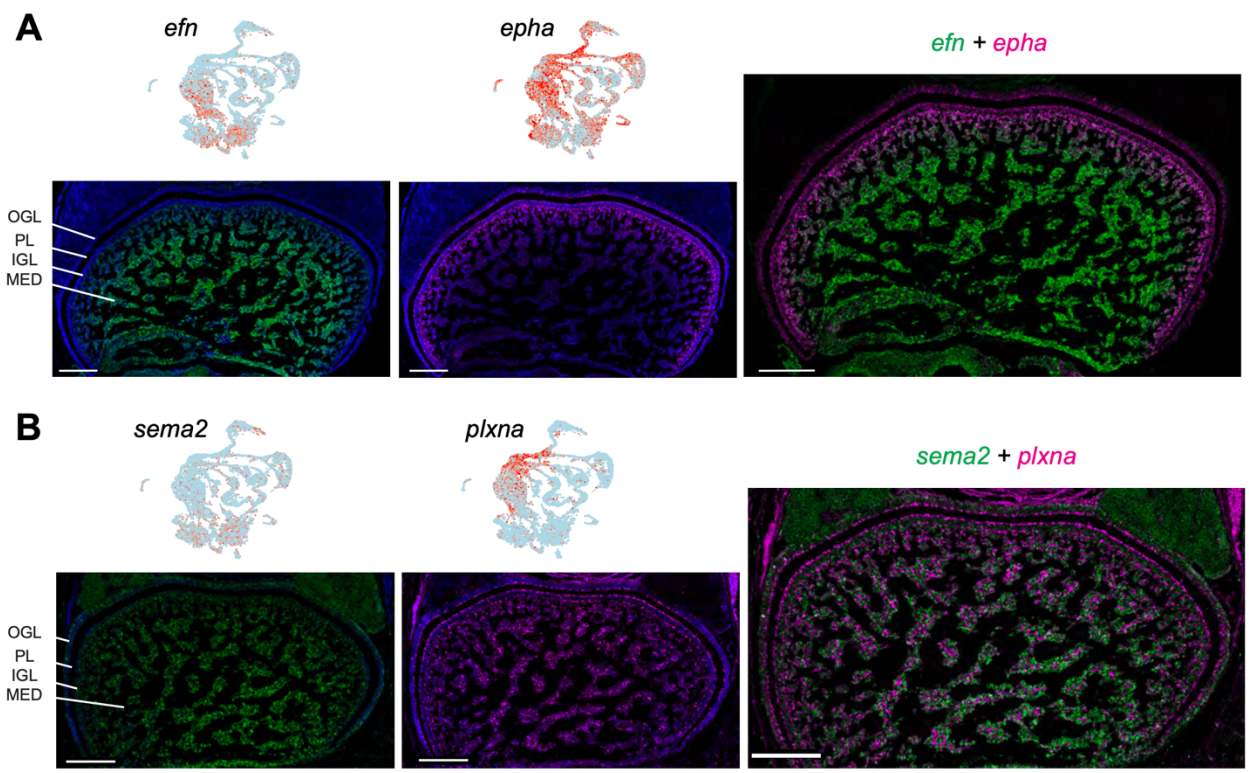

Figure 2.6: Expression of conserved patterning molecules
(A) Feature plots showing complementary scRNA-seq expression of an ephrin/Eph-receptor pair. FISH for these genes shows corresponding gradients of expression.
(B) Feature plots showing scRNA-seq complementary expression of a semaphorin/plexin pair. FISH for these genes shows corresponding spatial patterns of expression. In A,B, nuclei are stained with DAPI in blue, and scale bars indicate 200 um.
See also Table S2.2

Together, these data demonstrate that these putative immature neurons are found in distinct anatomical locations with discrete subtypes that can be molecularly defined. Moreover, conserved families of developmental genes, as well as novel cephalopod- or octopus-specific genes, define these subtypes, suggesting the possibility of both evolutionarily conserved and lineage-specific molecular mechanisms for development and function.

Cell-type and sub-layer organization of mature neurons in the optic lobe

A driving goal of this project was to identify the "parts list" of the octopus optic lobe and create an integrated model of cell type organization within their visual system. We therefore incorporated the findings for the mature neurons into a schematic spatial map of the optic lobe (Figure 2.7A-B). Here, we summarize this organization, and present multiplexed FISH data for markers within each layer to explicitly demonstrate the sub-layer organization.

Outer granular layer

We found four broad groups of cell types within the OGL (Figure 2.7C). First, a cluster of glutamatergic cells lines the most superficial aspect of the OGL, which also expresses hh. Second, one group of neurons from the dopaminergic+glutamatergic clusters is located in the central OGL, marked by an uncharacterized gene *obimac0010569*. Third, we identified a specific subset of octopaminergic neurons that co-expresses *tyrbh* and *pcdh-ob09200*. Finally, a diverse group of dopaminergic-only neurons spans sublayers of the OGL. This group falls into two major divisions: *six4/5*+ in the central OGL and dscam+ in the deep OGL. Notably, the expression of neuropeptides within the *dat*+ group also shows a progression across the depth of the OGL (Figure 2.7D).

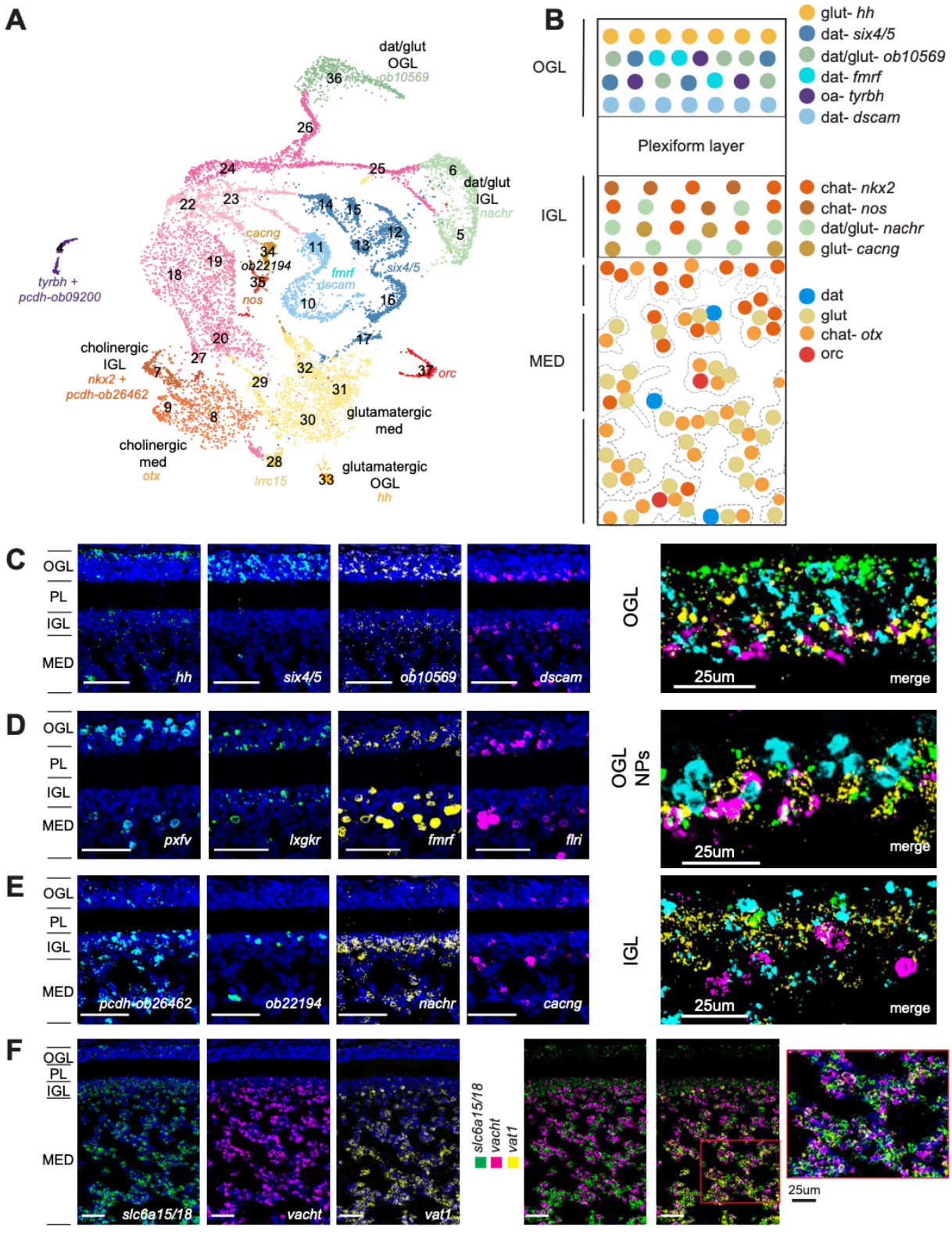

Figure 2.7: Summary of mature neuronal architecture

(A) UMAP showing cell subtypes in each neuronal class, along with annotation of spatial localization within the optic lobe.
(B) Schematic of cell type organization of the optic lobe. Genes defining subtypes are color-coordinated to match their clusters in (A).
(C) FISH showing sublayers of the OGL. Sublayers of OGL are demarcated from most superficial to deepest in order by expression of *hh, six4/5,* unidentified gene *obimac0010569*, and *dscam*. Throughout this figure, nuclei are stained with DAPI in blue, and, unless otherwise noted, scale bars represent 50um.

Inner granular layer

We found significant cell type diversity within the IGL (Figure 2.7E), consisting of at least four neuronal cell types with distinct sub-layer expression patterns. The largest population of cells

therein consists of neurons from one of the dopaminergic+glutamatergic clusters, which forms a distinct band as identified by expression of an *nachr*. A small population of cholinergic neurons expressing nos lines the superficial IGL (*obimac0022194*+ is shown in Figure 2.7E as proxy for *nos*+ cells), while a sub-group of glutamatergic neurons (cacng) lines the deep IGL. In addition, the *nkx2*+ cholinergic group spans from the IGL into the medulla (Figure 2.7E includes *obimac0026462* in place of *nkx2*).

Medulla

The medulla comprises a majority of the octopus optic lobe, and largely consists of two distinct populations of glutamatergic and cholinergic neurons (Figure 2.7F), with glutamatergic neurons more prevalent in the central tangential region. Notably, these are intermingled within the cell body “islands” of the medulla, suggesting that, at least at the level of these two large populations, there is no functional segregation across the islands. However, there are apparent distinctions in gene expression across the depth of the medulla. Within the cholinergic group, *nkx2*+ neurons are more superficial, while *otx*+ neurons are located throughout the medulla

(Figure 2.7B). In addition, there is a superficial-to-deep gradient of *vat1* expression that is shared across the cholinergic and glutamatergic neurons (Figure 2.7F). Finally, there are sparse populations of neurons that express extremely high levels of neuropeptides (Figure 2.4A, E). It is possible that these represent neurons projecting to downstream brain regions where these neuropeptides could play a role in regulating behavioral outputs[12,16,18,19,56].

Discussion

This study provides a comprehensive molecular description of neural subtype organization of the optic lobe cell types, to complement the anatomical descriptions provided by Young[5]. In addition to identifying non-neuronal and developing cells in our dataset, we reveal six major cell classes of mature neurons and a number of subtypes within these that correlate with discrete locations of the optic lobe, uncovering previously unknown cell type diversity and sub-layer organization of the octopus visual system. This study thereby contributes to a recently growing literature on transcriptomics of the cephalopod nervous system[57–59] and lays the basis for both the investigation of the role of distinct cell types in visual processing as well as the development of tools for targeting specific cell classes based on their molecular signatures. In addition to revealing the overall molecular architecture of the optic lobe, a number of the specific findings shown here have implications for elucidating the functional and development of the octopus visual system.

We found a wide array of cell types within the OGL, which was previously shown to consist of amacrine cells. Among the three broad classes of amacrine cells in the OGL that Young described[5], we found at least eight clusters of specific cell types that all have distinct spatial localizations within sublayers of the OGL (Figure 2.7C-D). These cell types are defined

by neurotransmitters (largely dopaminergic but also some glutamatergic), neuropeptides, a transcription factor (*six4/5*), an adhesion molecule (*dscam*), and a developmental signaling molecule (*hh*). This array of cell types bears a strong resemblance to the diversity of amacrine cells in the vertebrate retina, where over 60 amacrine cell types have been identified in mice[60]. However, in contrast to vertebrates, where amacrine cells primarily express the inhibitory neurotransmitters GABA or glycine[61], in the octopus we find that OGL neurons are predominantly dopaminergic. Notably, there is a specific population of dopaminergic amacrine cells in the vertebrate retina[62], and it has been shown that dendritic tiling in these amacrine cells is dependent on *dscam*[46], which strikingly is also expressed in a subset of dopaminergic OGL cells here. Moreover, in the vertebrate retina, distinct amacrine cell classes have been linked to a range of specific visual computations[63]. It will be intriguing to see whether similar functions can be assigned to the diversity of cells within the OGL based on the markers we have identified.

We identified a sparse but highly distinct population of neurons in the OGL that express *tyrbh*, the synthetic enzyme for octopamine. In both flies and mice, locomotion and arousal have profound effects on visual processing[64–66], which, in flies, are mediated by octopamine[67], and in mice are mediated in part by norepinephrine[6850]. Strikingly similar impacts of arousal on visual responses in octopuses were observed in an early study in the octopus EEG[69], suggesting a potential similar role for this octopaminergic system.

Another notable finding is the delineation of a large population of putative immature neurons within the octopus visual system. *O. bimaculoides* are born capable of feeding and living independently from hatching, relying on a variety of visually guided behaviors[27,28]. Despite this, octopuses continue to grow exponentially in size throughout their lifetime[70], increasing their mass by approximately 10x every two months, including growth of the optic lobes and other

brain regions[71,72]. Hence, the finding of such a large population of immature neurons is not surprising given the octopus's need for neurogenesis in order to support such massive brain growth. It has recently been shown in octopus embryos that neurogenesis occurs outside of the optic lobe, followed by long-distance migration into the optic lobe[73]. Therefore the immature population we observe likely represents post-mitotic, developing neurons that have recently completed migration, rather than a neurogenic population within the optic lobe itself.

In fish and birds, the visual system continues to grow throughout the animal's lifetime by expanding along the periphery in a proliferative marginal zone[74]. By contrast, we find that the immature neuronal population in the octopus is broadly distributed tangentially across the optic lobe, though with potential radial stratification along the borders of the OGL and IGL. Future work examining the ongoing developmental expansion of the optic lobe will be needed to reveal how new neurons coordinate their integration into the fully functioning visual system of the growing octopus.

We found genes associated with both the immature and mature cell types that may contribute to developmental establishment of identity, connectivity, and function, including both conserved (e.g. *hh, dscam, nkx, bib, pcdhs*) and novel (*obimac0011980, obimac0010569*) genes. We also identified subtypes within immature cell populations that are associated with mature neuronal cell types, suggesting different developmental trajectories and progenitor populations. Another striking developmental finding is the presence of complementary expression of ephrin/Eph-receptor and semaphorin/plexin pairs in the optic lobe, suggesting that these receptor-ligand pairs may play a similar role in setting up the spatial organization of the visual system as they do in flies and vertebrates.

Implications for future studies

We note that although we have delineated a number of subtypes within the major populations of cells in the octopus optic lobe, there is almost certainly additional diversity to be explored in future studies. Indeed, studies of the vertebrate retina have progressed from the early delineation of five major neuronal cell types to our current understanding of the tremendous diversity within each of these, including 30+ cell types within the retinal ganglion cells alone[75]. This initial description of cell types for the octopus visual system that we provide here can serve as a basis for delving further into such diversity.

In the current study, we defined cell types in the octopus visual system based on gene expression and related these cell types to their anatomical location. Future studies that relate these cell types to other aspects of neural identity, including anatomical morphology via single-cell labeling, downstream projection targets based on retrograde tracing, and visual response properties identified through calcium imaging, will be crucial in decoding cephalopod visual function. The molecular mapping we present here provides a roadmap for such studies and more generally provides a path forward towards cracking the functional, developmental, and evolutionary logic of the cephalopod visual system.

Methods

Experimental model and subject details

Octopus bimaculoides were obtained from the Cephalopod Resource Center at the Marine Biology Laboratory (Woods Hole, MA) and from Aquatic Research Consultants (San Pedro, CA). They were kept at the University of Oregon in a closed circulating 250 gallon aquarium system in artificial seawater, and fed daily on a rotating diet of frozen shrimp, crabs,

and fish. All husbandry and experimental protocols were in accordance with the EU 2010/63/EU[76] and AAALAC guidelines for the use and care of cephalopods for research.

Method details

Genome sample collection and sequencing

Optic lobe tissue was dissected from an adult female O. bimaculoides for whole genome sequencing. Tissue was sent to the University of Oregon Genomics & Cell Characterization Core Facility (GC3F) for DNA extraction and sequencing. High molecular weight genomic DNA was extracted using a Nanobind Tissue Big DNA kit (Circulomics). A Pacific Biosciences standard HiFi library was prepped with a SMRTbell Express Template Prep Kit 2.0. Genomic DNA was sheared at 20kb target size with a Megaruptor 2 instrument (Diagenode). BluePippin size selection (Sage Science) was used to omit the smallest fragments (<10-14kb) to enrich for longer fragments. Two HiFi genomic circular consensus sequencing (CCS) SMRTbell libraries were prepared as input for five HiFi SMRT cells. Single molecule sequencing of both libraries was conducted with a PacBio Sequel II system. After sequencing, data was imported into SMRT Link to generate 5.8 million HiFi reads with the CCS algorithm and create fastq files.

Four tissues were dissected from 6 week old juvenile octopuses to be used in Iso-Seq sequencing: optic lobe, central brain, retina, and arm. RNA extractions were performed using a RNeasy Plus Mini Kit (QIAGEN). A single bulk non-barcoded SMRTbell Iso-Seq library was prepared according to the manufacturer's protocol (PacBio) by GC3F. A multiplexed Iso-Seq library was sequenced across a single PacBio Sequel II SMRT cell. IsoSeq3 in SMRT Link was used to generate fastq files containing 1.08 million full-length transcripts.

Genome re-assembly and annotation

We used HiFiASM v0.15.5-r352[77] to assemble a contig-level genome with HiFi reads as input and default parameters. After initial assembly, duplications were removed using Purge_dups[78]. Protein-coding genes were annotated using existing bulk RNA sequences and our newly generated Iso-Seq data. Bulk RNA data was aligned to the genome assembly using Hisat v2.2.1[79] and gene predictions were assembled with StringTie v2.1.6[80] using parameters -c 4 -m 200 -j 3. All other parameters were set to default. To generate gene predictions with Iso-Seq data, we mapped the full length, non-chimeric reads (FLNC) to the genome using minimap2[81] using parameters -ax splice -uf --secondary=no -C5 -O6,24 -B4. After alignment, cDNA_cupcake (https://github.com/Magdoll/cDNA_Cupcake) was used to collapse alignments into transcript models. Unique transcripts with degraded 5' ends were filtered out of the final annotation file with filter_away_subset.py. StringTie and cDNA_cupcake annotations were combined using TAMA merge[82] with parameters -e longest_ends -d merge_dup.

The transcripts of the resulting gtf were used to run blastp against a Uniprot database and to run hmmer against the pfam database. The resulting hits were used as input for Transdecoder v5.5.0 (https://github.com/TransDecoder/TransDecoder) to predict single best coding regions. The existing mitochondrial genome and annotation were concatenated to the assembled genome and annotation files, respectively. This resulted in a final number of 18,896 gene annotations.

Orthologous relationships between our predicted genes and those of distant species were identified using OrthoFinder v2.5.2[31]. We used default parameters to cluster sequences into orthologue groups using sequences from eight species including *Homo sapiens* (hg38), *Mus musculus* (GCA_000001635.9), *Drosophila melanogaster* (GCA_000001215.4), *Aplysia californica* (GCA_000002075.2), *Crassostrea gigas* (GCA_902806645.1), Octopus sinensis

(GCA_006345805.1), and *Sepia pharaonis* (GCA_903632075.3). For orthologue groups that contained fewer than four genes, a tree was not generated. These genes were manually annotated using NCBI BLAST[32,83] to assign putative identity based on homology to deposited sequences in other species. Identities for putative neuropeptides (flri, fmrf, lxgkr, and pxfv) were assigned based on repeats within their predicted protein sequences.

Cell dissociation for scRNA-seq

Animals used for scRNA-seq were 6 week old juveniles with mantle lengths of 6.5mm-9.0mm. O. bimaculoides optic lobes were dissected on ice in Leibovitz-15 medium (Gibco) supplemented with 400mM NaCl, 10mM KCl, 15mM Hepes, 200 U/mL penicillin, and 0.2 mg/mL streptomycin. Single cell dissociation was performed by incubating tissue in papain (1 mg/ml; Worthington Biochemical Co) plus 1% DNase (10mg/ml in HBSS) in supplemented L-15 medium for 10 min at RT. The cells/tissue were gently pipetted up and down several times to dissociate large chunks. The cells/tissue were incubated for another 10 min at RT, pipetted up and down several times, and quenched in wash solution containing 2.5M glucose, 5mM Hepes, and 5% FBS in CMFSS (12mM Hepes, 435mM NaCl, 10.7mM KCl, 21mM Na2HPO4, 16.6mM glucose). Dissociated cells were passed through a 40 μM cell strainer (Fisherbrand), washed again, and resuspended in L-15 medium. A final sample cell concentration of 2000 cells per microliter, as determined on a BioRad TC20 cell counter, was used for cDNA library preparation. Dissociated samples were prepared in tandem, on the same day.

Single-cell cDNA library preparation

Sample preparation for two biological replicates was performed by the University of Oregon Genomics and Cell Characterization core facility (https://gc3f.uoregon.edu/). Dissociated cells were run on a 10X Chromium platform using 10x v.3 chemistry targeting 10,000 cells. The resulting cDNA libraries were amplified with 11 cycles of PCR and sequenced on either an Illumina Hi-seq or an Illumina Next-seq.

RNA Fluorescence In Situ Hybridization

Tissue collection for RNA fluorescence in situ hybridization consisted of juvenile octopuses (~4-6 weeks in age, 7mm in mantle length), which were first anesthetized in 4% EtOH in Artificial Seawater prior to fixation. Anesthetic replaced the seawater in the octopus' home chamber, and the chamber was placed on ice until the octopus was no longer ventilating or responsive. The mantle and arms were removed, leaving the central brain complex which was immediately placed into 10% Neutral Buffered Formalin. The brains were fixed for 24 hours at room temperature before being processed and embedded in paraffin and sectioned into 7um slices.

Custom probes were designed and ordered through Advanced Cell Diagnostics (ACDBio) (Hayward, CA). We followed the protocol available for ACDBio RNAscope[84], with minor changes to optimize it for use in paraffin-embedded octopus tissue. Briefly, we first removed the paraffin through baking, xylenes, and ethanol washes. We then fixed the tissue for 30 minutes in formalin at room temperature before dehydrating the tissue with an ethanol series. We proceeded with hybridization and target retrieval: 10min pre-treatment of H_2O_2, 12min target retrieval in a pressure cooker, and protease plus for 25min at 40C. Slides incubated with probes

for 2 hours before going into washes and 5X SSC overnight. On Day 2, we proceeded with amplification and used the appropriate HRPs and opal dyes before adding the HRP block. For multiple probes, additional HRP conjugates were added in a series-wise manner (HRP, opal dyes, block) before slides were mounted with DAPI and ProLong Gold Antifade.

Microscopy

Slides were imaged on the Leica SP8 confocal at 40x. Confocal images were scanned in a z-stack at 1um steps (2um steps for Figure S3) and were tiled. The resulting tiling merged image was then processed in FIJI[85]. The maximum intensity projection was generally taken across 8 planes (13 for Figure 1, 9 for Figure 6, 5 for Figure S3). When applicable, background subtraction was applied with a rolling ball radius of 100 pixels.

Quantification and statistical analysis

Cluster analysis

The sequencing data were analyzed using the 10X Cell Ranger pipeline, version 3.1.0 (Zheng et al., 2017) and the Seurat (version 3.1.4; Satija et al., 2015) software package for R, version 4.1.2, using standard quality control, normalization, and analysis steps.

Briefly, raw data from each biological replicate were read into R, with a minimum threshold of 3 cells and 500 genes. After visualizing the raw reads, genes, and mitochondrial percentage, we set thresholds of counts between 1000 and 20000, features above 600 genes, and less than 6 percent of mitochondrial content for downstream processing.

To correct for batch effects between our replicates, we followed guidelines for normalization and integration provided by [86] and [29] respectively. We selected integration features

within each dataset and applied SCTransform normalization before integrating the datasets based on Canonical Correlation Analysis[29]. For generating cell type clusters, we ran all of the following analyses on the "integrated" assay, but performed differential expression analysis on the "SCT" assay. Following standard downstream processing steps, we ran principal component analysis and UMAP on 25 dimensions. We ran FindNeighbors (dims 1:25) and FindClusters (resolution 0.85). We then generated a dendrogram and renumbered clusters based on this output. We identified the top differentially expressed markers and used these data to identify and subset the putative neurons. We also excluded one cluster from the rest of the analyses due to low number of transcripts and genes, suggesting this cluster did not represent real cellular expression. We re-ran UMAP on the subset of neurons and used this output for visualization and further cell type identification based on top differentially expressed markers. All UMAPs, including feature plots, are shown with datasets that are downsampled to 500 cells for visualization purposes. However, all dot plots show gene expression for full datasets. Feature plots for Figure 2B are shown with a min.cutoff of 0 and a max.cutoff of 1, except for *syt* which has a min.cutoff of 1 and a max.cutoff of 2. Feature plots for Figures 3-4 are shown with a min.cutoff of 0 and a max.cutoff of 4. Feature plots for Figures 5-6 are shown with a min.cutoff of 0 and a max.cutoff of 2.

References

1. Williamson, R., and Chrachri, A. (2004). Cephalopod neural networks. Neurosignals 13, 87–98.

2. Budelmann, B.U. (1995). The cephalopod nervous system: What evolution has made of the molluscan design. Experientia Supplementum, 115–138. 10.1007/978-3-0348-9219-3_7.

3. Nixon, M., and Young, J.Z. (2003). The Brains and Lives of Cephalopods (Oxford University Press).

4. Young, J.Z. (1961). Learning and discrimination in the octopus. Biol. Rev. Camb. Philos. Soc. 36, 32–96.

5. Young, J.Z. (1962). The optic lobes of Octopus vulgaris. Philos. Trans. R. Soc. Lond. B Biol. Sci. 245, 19–58.

6. Wells, M.J. (1962). Brain and Behaviour in Cephalopods (Stanford University Press).

7. Chiao, C.C., and Hanlon, R.T. (2001). Cuttlefish camouflage: visual perception of size, contrast and number of white squares on artificial checkerboard substrata initiates disruptive coloration. J. Exp. Biol. 204, 2119–2125.

8. Hanlon, R.T., and Messenger, J.B. (2018). Cephalopod Behaviour (Cambridge University Press).

9. Ogura, A., Ikeo, K., and Gojobori, T. (2004). Comparative analysis of gene expression for convergent evolution of camera eye between octopus and human. Genome Res. 14, 1555–1561.

10. Young, J.Z. (1962). The Retina of Cephalopods and Its Degeneration After Optic Nerve Section. Philosophical Transactions of the Royal Society of London B 245, 1–18.

11. Yamamoto, T., Tasaki, K., Sugawara, Y., and Tonosaki, A. (1965). Fine structure of the octopus retina. J. Cell Biol. 25, 345–359.

12. Young, J.Z. (1971). The Anatomy of the Nervous System of Octopus Vulgaris (Oxford University Press, USA).

13. Messenger, J.B. (1981). Comparative Physiology of Vision in Molluscs. Comparative Physiology and Evolution of Vision in Invertebrates, 93–200. 10.1007/978-3-642-67868-4_2.

14. Bullock, T.H., and Budelmann, B.U. (1991). Sensory evoked potentials in unanesthetized unrestrained cuttlefish: a new preparation for brain physiology in cephalopods. J. Comp. Physiol. A 168, 141–150.

15. Young, J.Z. (1979). The Nervous System of Loligo: V. The vertical lobe complex. Philosophical Transactions of the Royal Society of London B. Biological Sciences 285, 311–354.

16. Chichery, R., and Chanelet, J. (1976). Motor and behavioral responses obtained by stimulation with chronic electrodes of the optic lobe of Sepia officinalis. Brain Res. 105, 525–532.

17. Shigeno, S., Andrews, P.L.R., Ponte, G., and Fiorito, G. (2018). Cephalopod Brains: An Overview of Current Knowledge to Facilitate Comparison With Vertebrates. Frontiers in Physiology 9. 10.3389/fphys.2018.00952.

18. Young, J.Z. (1974). The central nervous system of Loligo. I. The optic lobe. Philos. Trans. R. Soc. Lond. B Biol. Sci. 267, 263–302.

19. Boycott, B.B., and Young, J.Z. (1961). The functional organization of the brain of the cuttlefish Sepia officinalis. Proceedings of the Royal Society of London. Series B. Biological Sciences 153, 503–534.

20. Saidel, W.M. (1982). Connections of the octopus optic lobe: an HRP study. J. Comp. Neurol. 206, 346–358.

21. Liu, Y.-C., Liu, T.-H., Su, C.-H., and Chiao, C.-C. (2017). Neural Organization of the Optic Lobe Changes Steadily from Late Embryonic Stage to Adulthood in Cuttlefish Sepia pharaonis. Front. Physiol. 8. 10.3389/fphys.2017.00538.

22. y Cajal, S.R. (1930). Contribución al conocimiento de la retina y centros ópticos de los cefalópodos (Unión Internacional de Ciencias Biológicas, Comité Español).

23. Konstantinides, N., Kapuralin, K., Fadil, C., Barboza, L., Satija, R., and Desplan, C. (2018). Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features. Cell 174, 622–635.e13.

24. Peng, Y.-R., Shekhar, K., Yan, W., Herrmann, D., Sappington, A., Bryman, G.S., van Zyl, T., Do, M.T.H., Regev, A., and Sanes, J.R. (2019). Molecular Classification and Comparative Taxonomics of Foveal and Peripheral Cells in Primate Retina. Cell 176, 1222–1237.e22.

25. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214.

26. Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. 19, 335–346.

27. Solorzano, Y., Viana, M.T., López, L.M., Correa, J.G., True, C.C., and Rosas, C. (2009). Response of newly hatched Octopus bimaculoides fed enriched Artemia salina: Growth performance, ontogeny of the digestive enzyme and tissue amino acid content. Aquaculture 289, 84–90.

28. Hanlon, R.T., and Forsythe, J.W. (1985). Advances in the laboratory culture of octopuses for biomedical research. Lab. Anim. Sci. 35, 33–40.

29. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21.

30. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049.

31. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238.

32. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

33. Deryckere, A., and Seuntjens, E. (2018). The Cephalopod Large Brain Enigma: Are Conserved Mechanisms of Stem Cell Expansion the Key? Front. Physiol. 9, 1160.

34. Messenger, J.B. (1996). Neurotransmitters of cephalopods. Invertebrate Neuroscience 2, 95–114. 10.1007/bf02214113.

35. Bröer, A., Tietze, N., Kowalczuk, S., Chubb, S., Munzinger, M., Bak, L.K., and Bröer, S. (2006). The orphan transporter v7-3 (slc6a15) is a Na -dependent neutral amino acid transporter (B0AT2). Biochemical Journal 393, 421–430. 10.1042/bj20051273.

36. Juorio, A.V., and Molinoff, P.B. (1971). Distribution of octopamine in nervous tissues of Octopus vulgaris. Br. J. Pharmacol. 43, 438P–439P.

37. Cornwell, C.J., Messenger, J.B., and Williamson, R. (1993). Distribution of GABA-like immunoreactivity in the octopus brain. Brain Res. 621, 353–357.

38. Seo, H.C., Curtiss, J., Mlodzik, M., and Fjose, A. (1999). Six class homeobox genes in drosophila belong to three distinct families and are involved in head development. Mech. Dev. 83, 127–139.

39. Kumar, J.P. (2009). The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. Cell. Mol. Life Sci. 66, 565–583.

40. Weasner, B.P., Anderson, J., and Kumar, J.P. (2004). The Eye Specification Network in Drosophila. Proc. Indian Natl. Sci. Acad. B70, 517–530.

41. Kawakami, K., Sato, S., Ozaki, H., and Ikeda, K. (2000). Six family genes--structure and

function as transcription factors and their roles in development. Bioessays 22, 616–626.

42. Krajniak, K.G. (2013). Invertebrate FMRFamide related peptides. Protein Pept. Lett. 20, 647–670.

43. Zatylny-Gaudin, C., and Favrel, P. (2014). Diversity of the RFamide Peptide Family in Mollusks. Front. Endocrinol. 5, 178.

44. Loi, P., Saunders, R., Young, D., and Tublitz, N. (1996). Peptidergic regulation of chromatophore function in the European cuttlefish Sepia officinalis. J. Exp. Biol. 199, 1177–1187.

45. Loi, P.K., and Tublitz, N.J. (2000). Roles of glutamate and FMRFamide-related peptides at the chromatophore neuromuscular junction in the cuttlefish, Sepia officinalis. J. Comp. Neurol. 420, 499–511.

46. Fuerst, P.G., Koizumi, A., Masland, R.H., and Burgess, R.W. (2008). Neurite arborization and mosaic spacing in the mouse retina require DSCAM. Nature 451, 470–474.

47. Yamagata, M., and Sanes, J.R. (2008). Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. Nature 451, 465–469.

48. Yang, C., Qi, Y., and Sun, Z. (2021). The Role of Sonic Hedgehog Pathway in the Development of the Central Nervous System and Aging-Related Neurodegenerative Diseases. Front Mol Biosci 8, 711710.

49. de Wit, J., Hong, W., Luo, L., and Ghosh, A. (2011). Role of leucine-rich repeat proteins in the development and function of neural circuits. Annu. Rev. Cell Dev. Biol. 27, 697–729.

50. Roeder, T. (1999). Octopamine in invertebrates. Prog. Neurobiol. 59, 533–561.

51. Pasterkamp, R.J. (2012). Getting neural circuits into shape with semaphorins. Nat. Rev. Neurosci. 13, 605–618.

52. Peek, S.L., Mah, K.M., and Weiner, J.A. (2017). Regulation of neural circuit formation by protocadherins. Cell. Mol. Life Sci. 74, 4133–4157.

53. Cifuentes-Diaz, C., Bitoun, M., Goudou, D., Seddiqi, N., Romero, N., Rieger, F., Perin, J.-P., and Alliel, P.M. (2004). Neuromuscular expression of the BTB/POZ and zinc finger protein myoneurin. Muscle Nerve 29, 59–65.

54. Rao, Y., Bodmer, R., Jan, L.Y., and Jan, Y.N. (1992). The big brain gene of Drosophila functions to control the number of neuronal precursors in the peripheral nervous system. Development 116, 31–40.

55. Wilkinson, D.G. (2001). Multiple roles of EPH receptors and ephrins in neural development. Nat. Rev. Neurosci. 2, 155–164.

56. Liu, T.-H., and Chiao, C.-C. (2017). Mosaic Organization of Body Pattern Control in the Optic Lobe of Squids. J. Neurosci. 37, 768–780.

57. Styfhals, R., Zolotarov, G., Hulselmans, G., Spanier, K.I., Poovathingal, S., Elagoz, A.M., Deryckere, A., Rajewsky, N., Ponte, G., Fiorito, G., et al. (2022). Cell type diversity in a developing octopus brain. bioRxiv, 2022.01.24.477459. 10.1101/2022.01.24.477459.

58. Duruz, J., Sprecher, M., Kaldun, J.C., Alsoudy, A., Tschanz-Lischer, H., van Geest, G., Nicholson, P., Bruggmann, R., and Sprecher, S.G. (2022). Molecular characterization of cell types in the squid Loligo vulgaris. bioRxiv, 2022.03.28.485983. 10.1101/2022.03.28.485983.

59. Gavriouchkina, D., Tan, Y., Ziadi-Künzli, F., Hasegawa, Y., Piovani, L., Zhang, L., Sugimoto, C., Luscombe, N., Marlétaz, F., and Rokhsar, D.S. (2022). A single-cell atlas of bobtail squid visual and nervous system highlights molecular principles of convergent evolution. bioRxiv, 2022.05.26.490366. 10.1101/2022.05.26.490366.

60. Yan, W., Laboulaye, M.A., Tran, N.M., Whitney, I.E., Benhar, I., and Sanes, J.R. (2020). Mouse Retinal Cell Atlas: Molecular Identification of over Sixty Amacrine Cell Types. J. Neurosci. 40, 5177–5195.

61. Diamond, J.S. (2017). Inhibitory Interneurons in the Retina: Types, Circuitry, and Function. Annu Rev Vis Sci 3, 1–24.

62. Witkovsky, P. (2004). Dopamine and retinal function. Doc. Ophthalmol. 108, 17–40.

63. Masland, R.H. (2012). The tasks of amacrine cells. Vis. Neurosci. 29, 3–9.

64. Niell, C.M., and Stryker, M.P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. Neuron 65, 472–479.

65. Chiappe, M.E., Seelig, J.D., Reiser, M.B., and Jayaraman, V. (2010). Walking modulates speed sensitivity in Drosophila motion vision. Curr. Biol. 20, 1470–1475.

66. Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J.A. (2015). Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. Neuron 86, 740–754.

67. Suver, M.P., Mamiya, A., and Dickinson, M.H. (2012). Octopamine neurons mediate flight-induced modulation of visual processing in Drosophila. Curr. Biol. 22, 2294–2302.

68. Polack, P.-O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state-dependent gain modulation in visual cortex. Nat. Neurosci. 16, 1331–1339.

69. Boycott, B.B., Lettvin, J.Y., Maturana, H.R., and Wall, P.D. (1965). Octopus optic responses. Exp. Neurol. 12, 247–256.

70. Forsythe, J.W., and Hanlon, R.T. (1988). Effect of temperature on laboratory growth, reproduction and life span of Octopus bimaculoides. Marine Biology 98, 369–379. 10.1007/bf00391113.

71. Yamazaki, A., Yoshida, M., and Uematsu, K. (2002). Post-hatching development of the brain in Octopus ocellatus. Zoolog. Sci. 19, 763–771.

72. Kerbl, A., Handschuh, S., Nödl, M.-T., Metscher, B., Walzl, M., and Wanninger, A. (2013). Micro-CT in cephalopod research: Investigating the internal anatomy of a sepiolid squid using a non-destructive technique with special focus on the ganglionic system. J. Exp. Mar. Bio. Ecol. 447, 140–148.

73. Deryckere, A., Styfhals, R., Elagoz, A.M., Maes, G.E., and Seuntjens, E. (2021). Identification of neural progenitor cells and their progeny reveals long distance migration in the developing octopus brain. Elife 10. 10.7554/eLife.69161.

74. Fernald, R.D. (1990). Teleost vision: seeing while growing. J. Exp. Zool. Suppl. 5, 167–180.

75. Sanes, J.R., and Masland, R.H. (2015). The types of retinal ganglion cells: current status and implications for neuronal classification. Annu. Rev. Neurosci. 38, 221–246.

76. Fiorito, G., Affuso, A., Basil, J., Cole, A., de Girolamo, P., D'Angelo, L., Dickel, L., Gestal, C., Grasso, F., Kuba, M., et al. (2015). Guidelines for the Care and Welfare of Cephalopods in Research -A consensus based on an initiative by CephRes, FELASA and the Boyd Group. Lab. Anim. 49, 1–90.

77. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods 18, 170–175.

78. Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics 36, 2896–2898.

79. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915.

80. Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 20, 278.

81. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100.

82. Kuo, R.I., Cheng, Y., Zhang, R., Brown, J.W.S., Smith, J., Archibald, A.L., and Burt, D.W. (2020). Illuminating the dark side of the human transcriptome with long read transcript sequencing. BMC Genomics 21, 751.

83. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

84. Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J., and Luo, Y. (2012). RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J. Mol. Diagn. 14, 22–29.

85. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682.

86. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 1–15.

Bridge

In Chapter II, I presented our work to define the cell types of the octopus visual system using single-cell RNA sequencing and multiplexed RNA fluorescence in situ hybridization experiments. In the early stages of this project, we attempted to conduct the scRNA-seq analyses with a highly fragmented genome assembly of the *O. bimaculoides* published in 2015. However, the low quality of the assembly limited the resolution of our single cell atlas. To address this issue, we decided to generate a de-novo genome assembly with long read sequencing—a type of sequencing that was unavailable in 2015. As a young graduate student, assembling and annotating a genome was no small task, however I successfully generated a high-quality, chromosome-scale octopus genome. This new assembly greatly improved the resolution of our single-cell atlas which enabled us to advance the work presented in Chapter II. Interestingly, we noticed some anomalies in the sequencing depth of one of the female octopus chromosomes. These anomalies led to an unexpected discovery and sparked an investigation into the sex determination system of extant cephalopods, which I explore in Chapter III of my dissertation.

III: CEPHALOPOD SEX DETERMINATION AND ITS ANCIENT EVOLUTIONARY ORIGIN

This chapter was published in the journal *Current Biology* in 2025. Co-authors include Dr. Silas Tittes, Dr. Scott T. Small, Dr. Jeremea O. Songco-Casey, Dr. Denise M. Piscopo, Dr. Judit R. Pungor, Dr. Adam C. Miller, Dr. Cristopher M. Niell, and Dr. Andrew D. Kern. G.C.C. and A.D.K designed the study. J.O.S.C., D.M.P., and J.R.P. prepared samples for sequencing. G.C.C. assembled the genome, G.C.C and S.T. annotated the genome, S.T. conducted the repeat analysis and generated the multiple species alignment, G.C.C. and S.T.S. conducted the synteny analyses. A.D.K did homology searches. G.C.C, S.T., S.T.S, and A.D.K wrote the manuscript. A.C.M, C.M.N, and A.D.K. helped supervise the project. All authors read and approved the final manuscript.

The citation for this publication is as follows:

Coffing, G.C., Tittes, S., Small, S.T., Songco-Casey, J.O., Piscopo, D.M., Pungor, J.R., Miller, A.C., Niell, C.M., and Kern, A.D. (2025). Cephalopod sex determination and its ancient evolutionary origin. *Current Biology*, 35(4), 931-939. https://doi.org/10.1016/j.cub.2025.01.005

Introduction

Octopuses, squids, and cuttlefishes – the coleoid cephalopods – are a remarkable branch in the tree of life whose members exhibit a repertoire of sophisticated behaviors[1]. As a clade, coleoids harbor an incredible variety of novel traits including the most complex nervous system among invertebrates, derived camera-type eyes, and rapid adaptive camouflage abilities[2,3]. The burst of evolutionary novelty that distinguishes cephalopods is even more striking in a phylogenetic context; cephalopods are a deeply diverged lineage that last share a common ancestor with other extant molluscs in the Cambrian period, roughly 550 million years ago[4,5]. With recent advances in genome sequencing technologies, we have the capability to explore the genomic foundations of cephalopod novelties. Here, using PacBio long-read sequencing of genomic DNA and IsoSeq full-length mRNA sequencing, we provide a novel chromosome-scale reference genome and annotation for a female California two-spot octopus (*O. bimaculoides*). Our assembly reveals that the female octopus has just one sex chromosome, consistent with a ZO karyotype, while the male has two (ZZ), providing the first evidence of genetic sex determination in cephalopods. We use our assembly and annotation in combination with existing genomic information from other cephalopods to create the first whole genome alignments from this group and demonstrate that the sex chromosome is of an ancient origin, before the radiation of extant cephalopods approximately 480 million years ago[4], and has been conserved to the present day in all cephalopod genomes available.

Results

A chromosome-level assembly reveals a hemizygous Z chromosome in female octopus

The California two-spot octopus was the first cephalopod genome to be sequenced in 2015[6] and subsequently placed into scaffolds in 2022[7]. Although these resources have been tremendously valuable for cephalopod research, the assembly still contains numerous gaps, and many gene annotations remain fragmented due to the highly repetitive nature of the genome. To sequence through the long, repetitive stretches of the *O. bimaculoides* genome, we re-sequenced a single female individual with PacBio's long, high-fidelity (HiFi) sequencing and used chromosomal conformation capture (Hi-C) to place scaffolds into chromosomes. After scaffolding, the total genome assembly size is 2.3Gb with 30 chromosomal scaffolds representing a N = 30 karyotype, which matches the findings of cytogenetic studies in other octopus species[8]. Additionally, we used IsoSeq sequencing to aid with annotating full-length genes. A comparison of assembly statistics to existing *O. bimaculoides* assemblies is shown in Table S3.1 and genome annotation statistics are in Table S3.2. Generally, our new assembly is more complete than previous assemblies, with a contig N50 of 0.86 Mb and a scaffold N50 of 101.05 Mb. The new assembly reduced the number of scaffolds in the assembly from 145,326[7] to 583, increasing the average scaffold length by roughly three orders of magnitude.

Strikingly, our Hi-C contact map showed evidence for one chromosome, chromosome 17, having reduced coverage in comparison to other chromosomes in our assembly from a female individual (Figures 3.1A & S3.1). As the original reference assembly from Albertin et al.[6] was from a male individual we were able to compare coverage in that assembly, which showed no

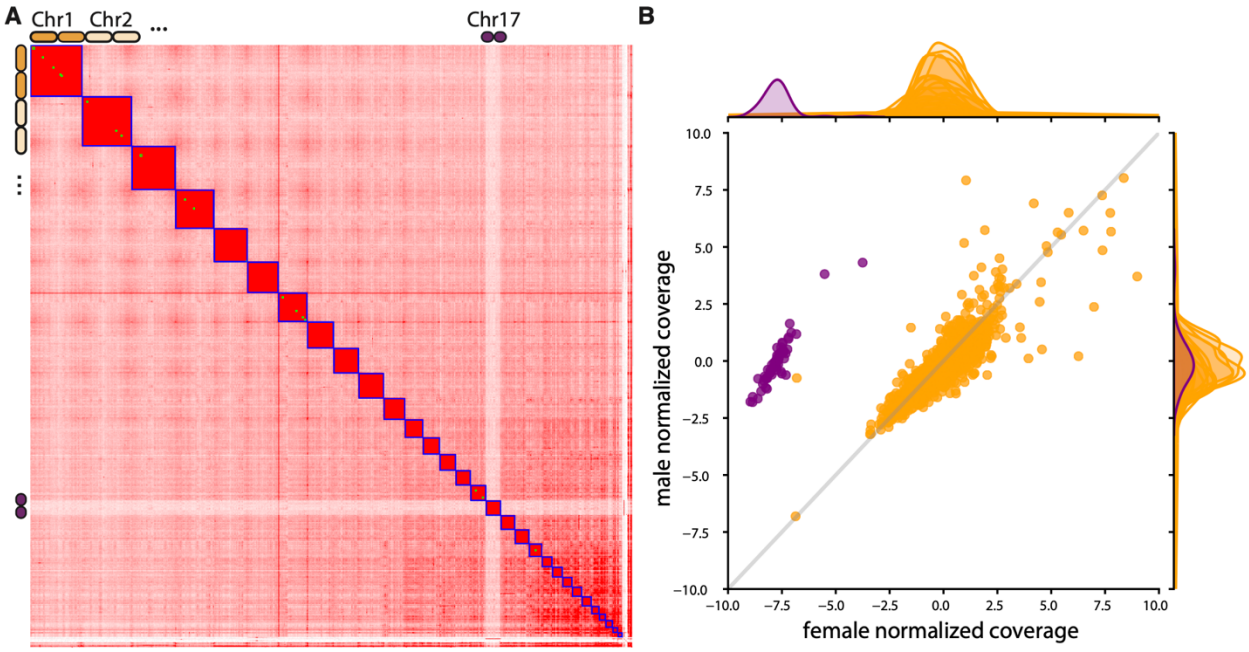

Figure 3.1: Sequencing data showing half coverage at chromosome 17 of *O. bimaculoides*.
(A) Hi-C contact map from our assembly of a female O. bimaculoides. Chromosome 17, our putative Z chromosome, is highlighted as a clear outlier in coverage.
(B) Normalized read depth of male and female whole-genome Illumina short read data. Purple points are from chromosome 17, whereas every other scaffold is shaded in orange. Scatterplot showing male vs. female normalized coverage, while the density estimates on the margins are from female (top) and male (side) separately. See also Figures S3.1 & S3.2 and Tables S3.1-S3.3.

difference in coverage between chromosome 17 and any other chromosome. Using short read Illumina data that we generated from unrelated female and male *O. bimaculoides* individuals (N = 2 of each sex; Table S2.3) we confirmed that females of this species are hemizygous for chromosome 17 whereas males are diploid, so hereafter we refer to chromosome 17 as Z (Figures 3.8B & S3.2).

Since the *O. bimaculoides* male genotype is clearly ZZ and the female is hemizygous at Z, we next turned our attention towards identification of a potential W chromosome limited to females. To do so we searched for scaffolds that were only present in the female-derived sequence libraries and absent from the male-derived libraries and assemblies. We found no such candidates, suggesting that females are ZO and males are ZZ. This perhaps indicates the

evolutionary loss of the W chromosome after substantial degradation[9,10]. It is possible that a small dot W exists that cannot be identified with our sequencing data. However, while ZZ/ZO sex determination systems are relatively rare among species found in the literature[11], they have been described in several groups including Lepidopterans[12], plants[10], amphibians[13], and more[14,15].

Genomic comparisons among cephalopods show the Z chromosome is an outlier

To determine if the Z chromosome was unique to *O. bimaculoides* or more broadly distributed among cephalopods, we inferred a species tree of cephalopods using protein sequences with the STAG algorithm in OrthoFinder[16]. This alignment compared three other octopus genomes (*Hapalochlaena maculosa*[17], *Octopus minor*[18], and *Octopus sinensis*[19]), two squid genomes (*Architeuthis dux*[20], and *Euprymna scolopes*[7]), a cuttlefish genome (*Sepia esculenta*), and the outgroup to coleoid cephalopods, the chambered nautilus (*Nautilus pompilius*[21]) (Figure 3.2A). Of 237,704 proteins input to OrthoFinder, 166,258 (69.9%) were assigned to 18,791 orthogroups. Next, we created the first whole-genome, multiple alignment among existing cephalopod genomes using the Progressive Cactus / Comparative Genomics toolkit pipeline[22] (see Supplemental Methods for details). Our alignment enables a host of analyses, including studies of sequence divergence and synteny, and we were particularly interested to compare patterns of evolution on autosomes versus the Z chromosome.

We calculated divergence in 1 megabase windows between *O. bimaculoides* and *O. sinensis* using the phast package (GTR model)[23], which revealed that the Z chromosome evolves significantly slower than autosomes (Mann–Whitney U test, $p = 0.0011$; Figure 3.2B). The same pattern holds true for Z chromosome divergence between *O. bimaculoides* and the Hawaiian

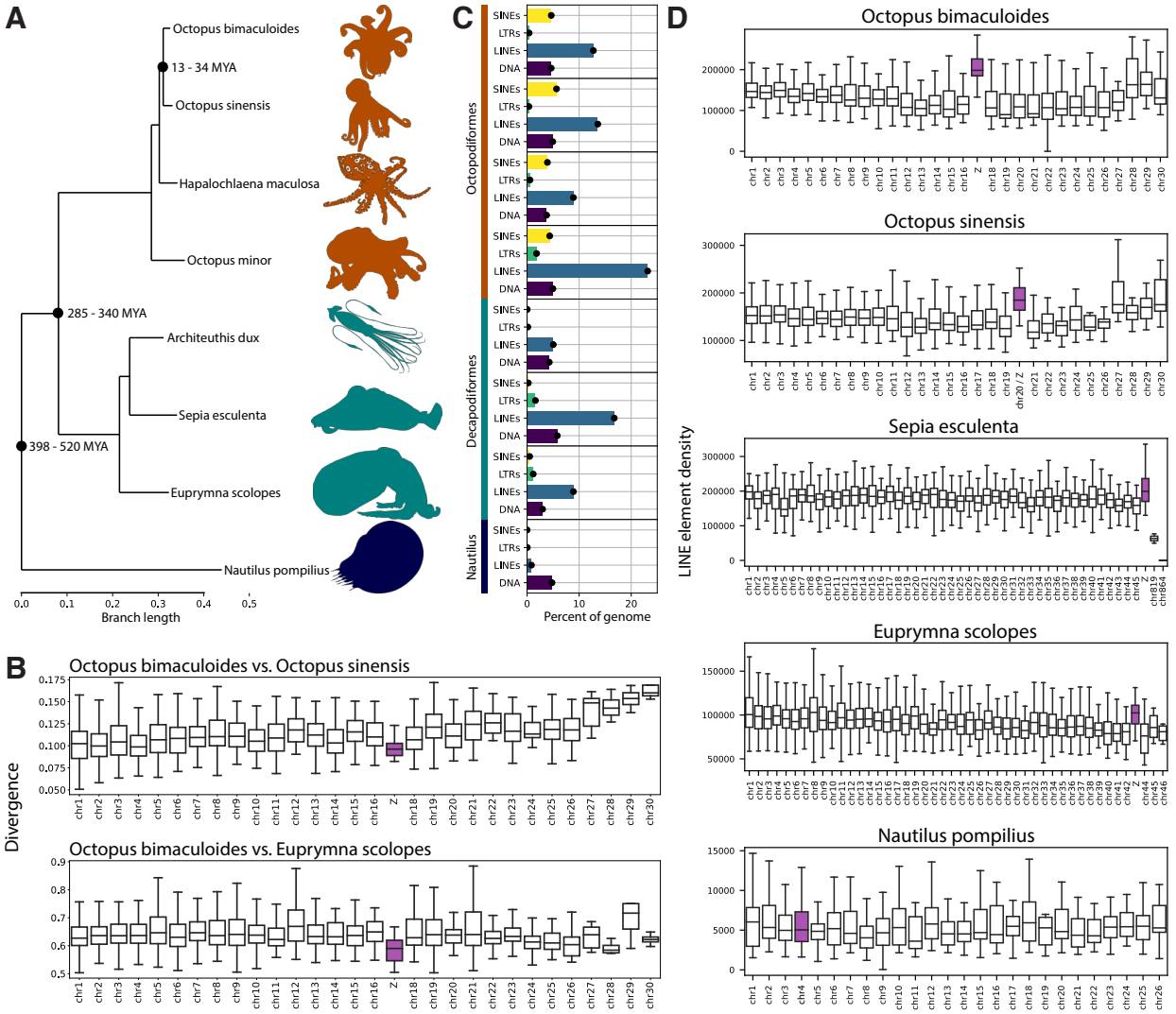

Figure 3.2: Repetitive element characteristics of the Z chromosome are unique

(A) Cephalopod phylogeny used in this study. Phylogeny inferred using one-to-one protein orthologs identified by OrthoFinder[16]. Branch lengths in units of amino acid substitutions per site. Divergence dates are from Huang et al.[4].

(B) Windowed pairwise divergence calculated between *O. bimaculoides* and *O. sinensis* (top panel) and *O. bimaculoides* and *E. scolopes* (bottom panel). Divergence was calculated in 1Mb, non-overlapping windows using our multiple genome alignment with *O. bimaculoides* as the reference. The Z chromosome is highlighted in purple.

(C) Genomic repeats associated with each leaf taxon. Repeats show percent of genome composed of SINE elements (yellow), LTR elements (green), LINE elements (blue), and DNA transposons (purple) for each species.

(D) LINE content across genomes of five cephalopod species. The number of LINE base pairs were counted in 1Mb windows. In each of the coleoid species, we found a single chromosome with elevated LINE density. These putative Z chromosomes have been highlighted in purple. In *N. pompilius*, no chromosomes have outlying LINE densities however the putative Z chromosome is highlighted in purple. See also Figures S3.3 & S3.4 and Table S3.5.

bobtail squid Euprymna scolopes (Mann–Whitney U tests, $p < 0.0001$; Figure 3.2B). We note that by using *O. bimaculoides* as a reference, the observed differences in divergence rates between Z and autosomes is likely conservative, as any loci which have moved between Z and autosomal chromosomes over evolutionary time should only obscure a difference, if present.

We also analyzed patterns of coding substitutions between *O. bimaculoides* and *O. sinensis*, restricting calculations to Z-linked genes in both species. While we found that the rate of nonsynonymous substitution was roughly equivalent between Z-linked and autosomal genes (Mann–Whitney U test, $p = 0.38$), Z-linked genes have decreased levels of synonymous substitution (mean Z-linked dS= 0.091, mean autosomal dS= 0.123; Mann–Whitney U test, $p < 0.00001$; Figure S3.4). Reduced divergence on the Z mirrors what is observed in the primate X chromosome relative to autosomes[24,25], and may be caused by a variety of factors including purging of recessive, mildly deleterious mutations from the hemizygous chromosome and sex-biased mutation rate differences between males and females. Using the ratio of the substitution rate at Z-linked versus autosomal loci allows us to estimate the degree of sex-biased mutation[35], generally termed α. We find that for synonymous divergence $\alpha = 0.127$ (95% C.I.= 0.022 − 0.268) meaning that for these positions we estimate the female mutation rate to be ∼7.8 times that of the male mutation rate. While synonymous substitutions are a decent proxy for neutral divergence, we can get a bulk estimate of α from the genome by simply considering divergence in large genomic windows. Using 1Mb windowed divergence yields an estimate of $\alpha = 0.539$ (95% C.I.= 0.406−0.703) indicating that the female mutation rate is ∼1.82 times that of the male mutation rate genome-wide. This observation of female-biased mutation is notable and has been seen in fish lineages[36] but contrasts to the commonly-observed male bias in amniote lineages[36,37]. Although female-biased mutation rate is quite rare, *O. bimaculoides* females undergo

synchronous ovulation[38], a process that has been hypothesized to contribute to higher mutation rate in females[36].

Repetitive element characteristics of the Z chromosome are unique

Our whole-genome alignment allowed us to examine repetitive element evolution in a phylogenetic framework. Figures 3.2A & 3.2C show a phylogenetic tree representing the relationship among these cephalopods along with the percent of the genome occupied by each of four repeat element classes: DNA transposons, LINEs, LTRs, and SINEs. It is clear from this comparison that the lineage leading to octopuses underwent a dramatic increase in the number of SINE elements, with SINEs taking up approximately 4.64% of the genome sequence in *O. bimaculoides*. The SINE expansion in octopus genomes has been noted previously with sparser comparisons[5], however when seen in a phylogenetic light (Figure 3.2) this pattern is abundantly clear.

While a historical expansion of SINE elements is a striking feature of the octopus genome, we also found an abundance of LINE elements, which composed a higher proportion of the genome than SINE elements, at 12.73%. Two clades of LINE elements, R2/R4/NeSL and RTE/Bov-B, made up the bulk of this with 3.10% and 4.59% of the genomic sequence respectively (Figure 3.2C).

Our chromosome-level assembly allowed us to explore if there is genomic heterogeneity in the accumulation of transposable elements (TEs). Visualization of the TE contributions to individual chromosomes (Figures 3.2D, S3.3) suggests that while most chromosomes have little variation in their proportions of TEs, the *O. bimaculoides* Z chromosome has an elevated density

of LINE elements, harboring LINEs at approximately twice the density as the genome-wide background (Mann-Whitney test; $p < 0.0001$; Figure 3.2D).

LINE elements are a signature of coleoid sex chromosomes

We hypothesized that the abundance of LINE elements could be a clear signature of Z chromosomes in other cephalopod genome assemblies. Looking at the landscape of TEs in the high quality, chromosome-level assemblies of *O. sinensis*, *E. scolopes,* and *S. esculenta* yielded a striking pattern: a single large chromosome in each of those assemblies (chrZ in *S. esculenta,* formerly chr20 in *O. sinensis* and formerly chr43 in *E. scolopes)* contained a high density of LINE elements, perhaps suggesting the presence of an orthologous Z chromosome (Figure 3.2D). As in *O. bimaculoides*, each of these chromosomes harbors a significantly greater density of LINEs in comparison to the rest of their genomes (Mann-Whitney tests; $p < 0.0001$, $p = 0.00023$ and $p < 0.0001$, for *O. sinensis*, *Euprymna scolopes*, and *Sepia esculenta* respectively; Table S3.5). As a result of these findings supporting our hypothesis, we renamed the putative Z chromosomes to Z in all figures for clarity. Upon inspection of the much more distant Nautilus genome, no chromosomes with elevated LINE element content could be identified. Thus, evidence from LINE element enrichment suggests that this could be a unique feature of the coleoid Z chromosome that may have originated before the split of the squid and octopus lineages, but after divergence from Nautilus.

Syntenic relationships of genes on the Z chromosome are conserved

We hypothesized that the chromosomes with elevated LINE densities in *O. sinensis*, *E. scolopes,* and *S. esculenta* might be orthologous to the *O. bimaculoides* Z chromosome. To test

this hypothesis, we compared synteny among chromosomes using gene-level annotations as revealed using the GENESPACE and MCScanX software packages[39–41], which use orthologous gene identification to define conserved synteny blocks. As seen in Figures 3.3A and 3.3B, gene-based synteny confirms our hypothesis – the *O. bimaculoides* Z chromosome has a large block of conserved synteny on the putative Z chromosomes in *O. sinensis*, *E. scolopes,* and *S. esculenta.* Dotplots showing finer resolution pairwise alignments are shown in Figure S3.5.

Due to the huge evolutionary distances between *Nautilus pompilius* and the other cephalopod taxa studied here, coupled with the extensive genome expansion along the coleoid lineage, calling syntentic regions based on colinear blocks of one-to-one orthologs across multiple taxa is challenging. Instead, we used MCscanX, a method not constrained to single copy genes, to compare chromosome Z of *O. bimaculoides* to the *N. pompilius* genome in a more targeted fashion. This revealed 1,057 collinear genes (8.31% of all genes) contained on 135 alignment blocks between *O. bimaculoides* Z-linked proteins and *N. pompilius*. Syntenic blocks are found across almost all the N. pompilius chromosomes, indicative of the extensive genome and gene duplication since coleoid cephalopods diverged from *N. pompilius* (Figure S3.6). Of the 300 single copy, unique Z-linked genes in *O. bimaculoides*, more than half were contained on one of two chromosomes in *N. pompilius*: 87 were contained in collinear blocks on chromosome 7 and 77 were found on *N. pompilius* chromosome 4. We note that neither of these chromosomes is unusual for LINE density (see Figure 3.2D).

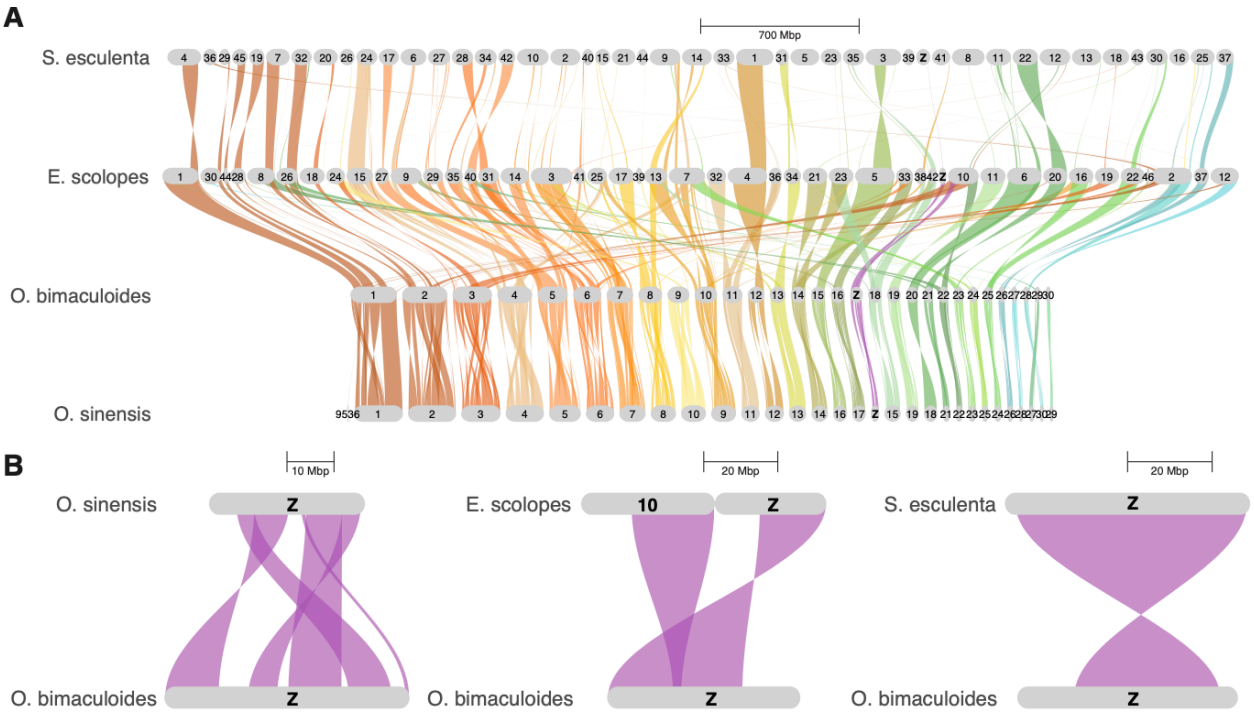

Figure 3.3: Syntenic relationships between four cephalopod species
(A) Riparian synteny plot generated using orthogroups between the chromosomes of representative octopus, squid, and cuttlefish species. Ribbons connected to the *O. bimaculoides* Z chromosome are highlighted in purple.
(B) Paired riparian synteny plots between the Z chromosome of *O. bimaculoides* and putative Z chromosomes of three other species, *O. sinensis, E. scolopes,* and *S. esculenta*, reveal conservation of the Z chromosome. See also Figures S3.5 & S3.6.

A single, ancient origin of the cephalopod Z chromosome

Our evidence, when taken together, demonstrates that the *O. bimaculoides* Z chromosome is a genomic outlier with clear homology at the gene level and with respect to repeat content to the putative Z chromosomes in *O. sinensis*, *E. scolopes,* and *S. esculenta.* While the divergence time between *O. bimaculoides* and *O. sinensis* is relatively modest, between 13 - 34 million years[1], divergence between Decapodiformes (squid and cuttlefish) and Octopodiformes (octopus) is much older, with an approximate divergence time between *O. bimaculoides* and *E. scolopes* of 285 - 340 million years[1]. Thus, it was imperative to examine whether the orthologous chromosomes in other cephalopod species are hemizygous. Figure 3.4

presents sequencing coverage data from four additional species, including short read libraries that we generated from sexed Octopus bimaculatus individuals (see methods), the sister taxa to *O. bimaculoides*, a female cuttlefish *S. esculenta* recently released by the Darwin Tree of Life project (NCBI, GCA_964036315.1), Illumina short read libraries of two unsexed individuals of N. pompilius[1,30] and a set of eight Illumina short read libraries derived from single, unsexed embryos of *E. scolopes*[10]. These results again confirmed our hypothesis in each case. In species where we have sexed individuals *(O. bimaculatus* and *S. esculenta), the putative* Z chromosome appears at approximately half coverage in the female in comparison to representative autosomes (chromosomes 1 and 10 shown as autosomal comparisons in Figures 3.4A, 3.4B; all chromosomes shown in S3.7, and S3.8). For *Euprymna scolopes* we did not have sexed individuals, however in the eight embryos coverage among chromosomes was nearly uniform across individuals except at the putative Z (chr43). The putative Z is an outlier in comparison to the other chromosomes with multiple distinct coverage classes among individuals, suggesting females are hemizygous as observed in *O. bimaculoides* (Figures 3.4C; all chromosomes shown in S3.9). Finally, in comparing sequencing coverage among the two available unsexed Nautilus genomes, we again observed a single chromosome, chr4, at reduced coverage in one of the two libraries (Figure 3.4D). Chr4 is one of two chromosomes in Nautilus that shows synteny to the *O. bimaculoides* Z as determined by our MCScanX analysis. Thus, while we were not able to detect a Z chromosome based on LINE element density in Nautilus, sequencing coverage among individuals and synteny strongly suggests that a sex chromosome is present and that it shows orthology to the Z in coleoid cephalopods.

Having established strong support for homology of the Z chromosome among cephalopods, we were lastly interested in examining what genes might be shared in orthologous

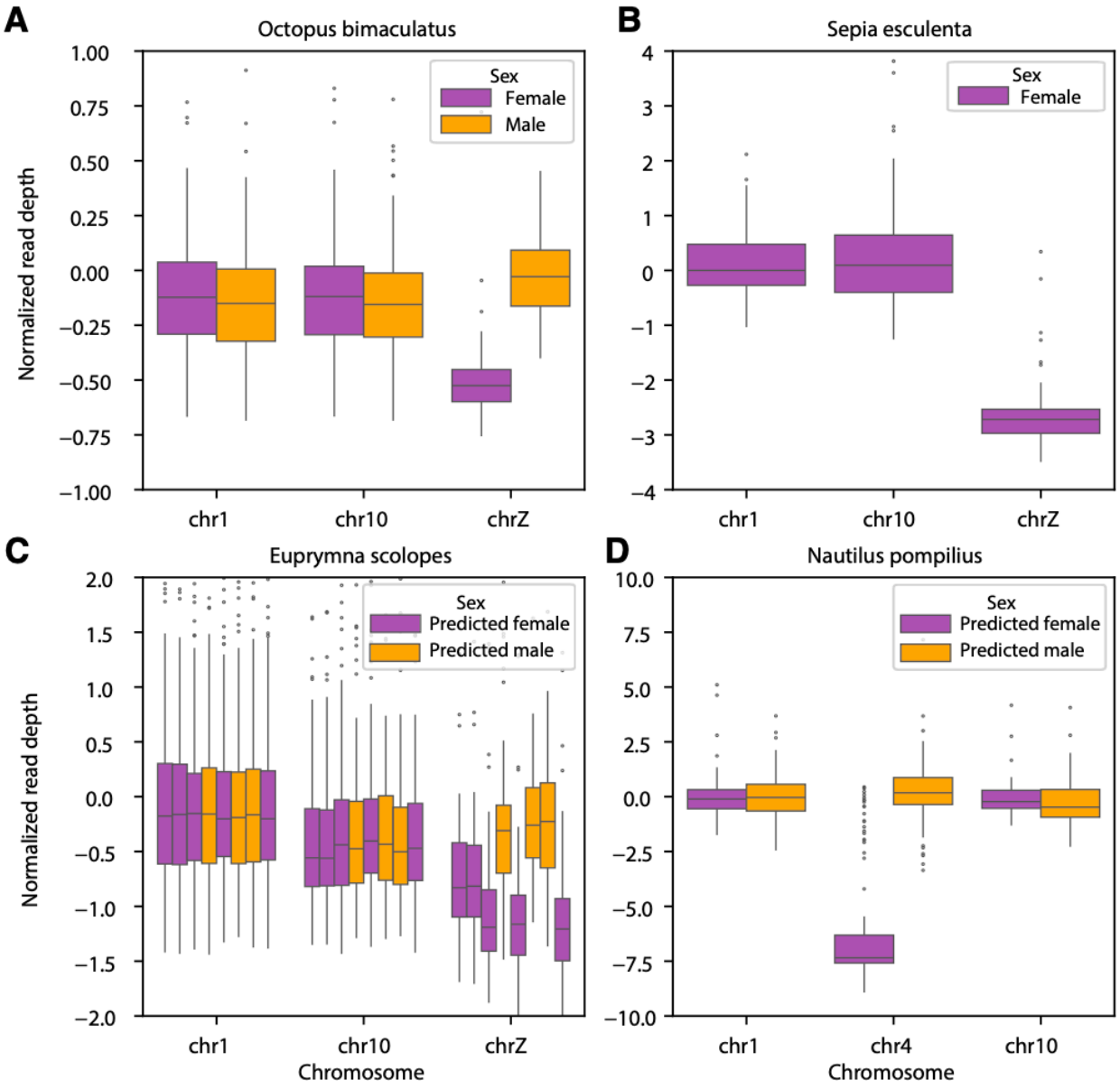

Figure 3.4: Normalized coverage of sequencing data mapped to representative chromosomes in *O. bimaculatus, S. esculenta, E. scolopes and N. pompillius.*

(A) For *O. bimaculatus*, we generated Illumina short reads from two male and two female individuals (See Methods for more details).

(B) Female *S. esculenta* short reads were generated by Darwin Tree of Life Project and downloaded from NCBI (GCA_64036315.1).

(C) Sequencing data for unsexed embryos of *E. scolopes* were generated by Schmidbaur et al.[10]. The putative Z chromosome is the only chromosome with variation in coverage across individuals. Predicted female individuals (purple) have significantly lower coverage at the presumed Z chromosome compared to the average coverage of all other chromosomes ($p < 0.0005$ in all cases; Mann Whitney U tests) and predicted males (orange) do not (N.S.; Mann-Whitney U tests;).

(D) In *N. pompilius*, Illumina short reads from the predicted female individual were generated by Zhang et al.[30] and Illumina short reads from the predicted male individual were generated by Huang et al.[1]. See also Figures S3.7-S3.9 and Table S3.4.

regions of the Z chromosome among lineages. We first compared annotations from the shared syntenic chromosome Z blocks defined by GENESPACE across coleoid cephalopods, using annotation from *O. bimaculoides, O. sinensis,* and *E. scolopes. This revealed* 19 unique protein coding loci shared among these three taxa that are housed in this region of the genome. We did blastp homology searches of these genes to human proteins and found 16 of 19 had strong hits (Table S3.6). Using publicly available summaries from GeneCards[42] we report that all 16 show mRNA expression in human reproductive tissues, and 15 of 16 of these show protein expression in human reproductive tissues. A particularly leading hit is the protein obimac0008950, which shows strong homology to the human Sperm Associated Antigen 9 (*SPAG9*; e-val=0.0). We next asked if the 19 proteins we identified above also occur in the colinear blocks identified between the Z in *O. bimaculoides* and *N. pompilius* and found that 9/19 were present, and 6 of these 9 could be found in humans. All 6 show evidence of mRNA expression and 5 of these 6 show protein expression in human reproductive tissues. Taken together, these blast results let us speculate that the genes retained on the Z chromosome may represent an ancient set of proteins essential to animal reproduction and/or gametogenesis. Last, we searched for known sex determination genes from other mollusc species on the Octopus Z. Four are described: Forkhead box protein L2 (*FOXL2*) in *M. yessoensis* and *C. farreri*, Zinc finger protein 226-like (*ZNK226l*) in *P. magellanicus,* and Cytochrom P450 3A24-like (*CYP3A24l*) in *A. japonicum* have been found to be female-biased[43]. Doublesex and mab-3 related transcript factor 1-like (*DMRT1L*) is mollusc-specific and male-biased in *M. edulis*[44]. We downloaded representative protein sequences of these genes from NCBI (*FOXL2* and *DMRT1L* from *A. irradians*, *ZNK226l* and *CYP3A24l* from *M. yessoensis*) and ran BlastP searches to a database composed of proteins we annotated on the *O. bimaculoides* Z chromosome. This revealed Z-linked hits for *FOXL2*

(obimac 0009054.1, e-val=3.23e-33) and *ZNK226l* (obimac_0008819.2, e-val=2.66e-27), thus the *O. bimaculoides* Z harbors known mollusc sex determination genes.

Discussion

Our results provide the first glimpse of sex determination in cephalopods, a phenomenon which until now has remained a mystery. The clear presence of a hemizygous Z chromosome in females suggests a ZZ/ZO sex determination system that had previously been missed in cytological comparisons. Since ovaries contain fewer cells in meiosis than testes, it is difficult to identify female heterogameity via karyotyping[20]. Therefore, it is standard to use testicular tissue[45,46] or unsexed embryos[47] for karyotyping. Although we have not found evidence for a W chromosome in *O. bimaculoides* or any other cephalopod species, assembly of degenerate chromosomes is particularly challenging and so evidence for a W might be gathered in the future. Taken together, our evidence suggests that the cephalopod Z chromosome evolved once in the lineage leading to the common ancestor of all extant lineages. Recent molecular estimates of the time to the common ancestor of Nautilus and coleoids place it at ~482 million years ago (range 398-520 Mya)[1] thus the Z chromosome would had to have originated at least this long ago. This is an astoundingly long time for a sex chromosome to be preserved[18]. A few other ancient sex chromosomes have been described previously from liverworts (~430Mya)[48] and mosses (~300Mya)[49], and there is evidence that the insect X chromosome is quite ancient (~450Mya)[50], although it shows rapid evolution in some clades. Thus in context, the cephalopod Z chromosome is among the very oldest sex chromosomes yet described.

Methods

Experimental model and study participant details

Sample collection

A description of the genome sample collection and sequencing was published previously[43]. Optic lobe tissue was collected from an adult female O. bimaculoides that was obtained from the southern coast of California. Optic lobe, central brain, retina, and arm tissue were dissected from related 6-week-old juvenile octopuses to be used for construction of an Iso-Seq library. Additional arm tissue was dissected to construct an Illumina library to be used in Hi-C sequencing. Four additional unrelated individuals were collected from the southern coast of California for sex chromosome coverage analysis. Optic lobe tissue was extracted from two individuals, gill tissue from one individual, and testes from one individual (Table S3.3). Optic lobe tissue of four unrelated Octopus bimaculatus individuals were also collected from the southern coast of California for sex chromosome coverage analysis (Table S3.4).

Method details

Library construction and sequencing

All tissue samples were sent to the University of Oregon Genomics & Cell Characterization Core Facility (GC3F) for DNA extraction, sample preparation, and sequencing. High molecular weight genomic DNA was extracted with a Nanobind Tissue Big DNA kit (Circuloimics) from the adult optic lobe tissue. A SMRTbell Express Template Prep Kit 2.0 was used to construct a Pacific Bioscience standard HiFi library. Megaruptor 2 instrument was used to shear genomic DNA at 20kb target size and BluePippin size selection (Sage Science) omitted the smallest fragments (<10-14kb). Two HiFi genomic circular consensus sequencing (CCS)

SMRTbell libraries were prepared as input for five HiFi SMRT cells. For Iso-Seq sequencing, a RNeasy Plus Mini Kit (QIAGEN) was used to conduct RNA extractions. GC3F prepared a multiplexed SMRTbell Iso-Seq library according to PacBio's protocol to be used as input for one SMRT cell. Single molecule sequencing of HiFi and Iso-Seq libraries was conducted with a Pacific Biosciences Sequel II system (Table S3.3). For scaffolding, we constructed one Illumina library with a Phase Hi-C kit to produce two NovaSeq 6000 sequencing Hi-C runs (Table S3.2). For the male vs. female coverage analysis, genomic DNA from each individual was barcoded with GC3F's TruSeq-style adaptors. The barcoded DNA was pooled and sequenced on two Illumina NovaSeq 6000 S4 lanes.

Post sequencing, the genomic HiFi reads were processed with SMRT Link to generate 5.8 million HiFi reads. The IsoSeq subreads were processed with the IsoSeq3 v.3.0.8 pipeline. Briefly, CCS generated HiFi reads. Primers were removed and barcodes were identified with lima. Reads were further refined by trimming poly(A) tails and removing concatemers. Finally, reads were clustered and collapsed into unique isoforms with parameters –min-alncoverage 0.75 –min-aln-identity 0.75.

Quantification and statistical analysis

Genome size estimation and assembly

The O. bimaculoides genomic HiFi reads were used to estimate genome size. We generated a k-mer frequency distribution with JellyFish v.2.3.0[44], which was input into GenomeScope v.2.0[45] to generate genome size estimation plots. Estimation with 21-mers estimated the genome to be 2.0Gb with 0.838% heterozygosity. Estimation with 31-mers estimated the genome to be 2.03Gb with 0.653% heterozygosity (Figure S10).

Initial genome assembly was conducted using HiFiasm v.0.15.5-r352[46] with HiFi reads and standard parameters. Hifiasm did not sufficiently remove duplicates, so purge dups v.1.2.5[47] was used to purge haplotigs and overlaps. We placed the resulting scaffolds into a chromosome-level assembly using the data generated from the Hi-C library. The SLURM-compatible scripts of Juicer v.1.6[48] were used to identify Hi-C contact points on the genome. 3D de novo assembly (3D-DNA) pipeline v.190716[49] was used to correct genome mis-assemblies, anchor order of chromosomes, and orient genomic fragments.

Genome annotation

Transposable elements were identified using Repeat Modeler[50] and were annotated with Repeat Masker v.4.1.5[51]. Protein-coding genes were annotated using both existing RNA-seq data and our newly generated Iso-seq data. Using existing RNA-seq databases, we ran Braker v.2.1.6[52] to annotate. Separately, we generated annotations with the Iso-seq reads using SQANTI3 v.5.1[53]. The full-length isoforms generated by IsoSeq3 were filtered with SQANTI3 filter using default parameters. After generating the RNA-seq and Iso-seq based annotation files, they were merged using TAMA merge[54] with parameters -e longest ends -d merge dup. Genome annotation statistics that were calculated with AGAT's agat_sq_stat_basic.pl script[55] are presented in Table S3.2.

Orthology inference

OrthoFinder v.2.5.4[16] was used with default parameters to cluster the protein sequences of O. bimaculoides into ortholog groups using sequences from seven additional cephalopod species including Hapalochlaena maculosa, Octopus minor, Octopus sinensis, Architeuthis dux,

Euprymna scolopes, Sepia pharaonis, and Nautilus pompilius. The rooted species tree was used as input for the multiple species genome alignment.

Multiple species alignment

We constructed whole genome alignments of eight cephalopod species with available reference genomes (Table S3.5) using progressive cactus v2.3.0 (via Docker container)[22]. We used the whole genome alignments and phastCons v.1.5[56] to identify highly conserved elements on each scaffold. This alignment, as well as our new reference genome and annotation are available as a UCSC genome assembly hub from http://poppy.uoregon.edu/~ssmall/ucsc_genome_hubs/v2.1.0/Octo.v2.1.0/hub.txt

Coverage calculations

To estimate read coverage of each species, we aligned short reads with bwa-mem2[57] and long reads with minimap2[58] to the respective chromosome-scale genomes. Sequencing depth was estimated using mosdepth[59] in windows of 500,000bp. Plotted coverages of all species were normalized by the median coverage of the first chromosome.

Annotation of Sepia esculenta

At the time of these analyses, a genome annotation of Sepia esculenta was not yet publicly available. In order to include Sepia esculenta into the synteny analysis we lifted over the genome annotation from Sepia pharaonis[60] onto Sepia esculenta using the software package Liftoff[61] with default settings.

Chromosome renaming

To simplify our analyses and plots we renamed all full-length chromosomes in all the genomes that we included in this manuscript. We renamed chromosomes from longest to shortest with the naming scheme chr1, chr2, chr3, etc. with the exception of putative sex chromosomes which were renamed as chrZ.

Assessing sequence divergence among scaffolds

We converted the hierarchical alignment (hal) generated by progressive cactus to maf with hal2maf. We filtered the full, repeat-masked alignment into 1Mb windows along the genome using the phast utility msa view, where windows were made with bedtools makewindows v2.31.0[62]. We further filtered alignments with multiple sequences per species using maftools mafDuplicateFilter v.0.1[63]. We used phyloFit to construct phylogenies on each resulting 1Mb alignment region, only using alignments that had less than 40% of positions with gaps using the GTR mutation model. We calculated divergence as the cophenetic distance between O. sinensis and O. bimaculoides using the cophenetic.phylo function from the R package ape v.5.7-1[64]. The ratio of male to female mutation rates, α, was estimated according to Miyata et al., 1987. Let R represent the ratio of Z-linked divergence to autosomal divergence. Then, $\alpha = \frac{3R-2}{4-3R} \cdot \alpha$ was estimated separately for mean dS at the invidual gene level as well as for mean divergence in 1Mb genomic windows. Confidence intervals on α were computed using bootstrap resampling of the observed data.

Synteny analyses

Synteny analyses between all chromosomes of O. bimaculoides, O. sinensis, E. scolopes and S. esculenta were conducted using the R package GENESPACE v.1.2.3[30]. O. bimaculoides was used as the reference species for all synteny maps. In all riparian plots and dotplots, chromosomes are scaled by physical position (useOrder = FALSE parameter in GENESPACE).

To identify the macrosynteny relationships between N pompilius and other cephalopod species, we performed a macrosynteny analysis using MCScanX[31]. Input files were prepared following protocols described in Wang et al.[32] and using protein sequences assigned to the Z chromosome in O. bimaculoides and 15,086 N pompilius proteins assigned to chromosomes[4]. MCScanX was then run with default parameters and results visualized using scripts included with MCScanX program (www.github.com/wyp1125/MCScanX).

dN and dS estimations

To estimate dN and dS, we used protein alignments of O. bimaculoides and O. sinensis. Proteins of 8,245 single copy orthologs previously identified using OrthoFinder v2.5.4[16], were re-aligned using MACSE v2.07[65,66], with -prog alignSequences and default options. Z and autosome alignments were separated based on previous synteny analyses. In O. sinensis, chr20 (original name NC043016) was found to be orthologous to O. bimaculoides chrZ. Therefore, only proteins that were single copy orthologs between the O. bimaculoides chrZ and the inferred O. sinensis chrZ were considered for the sex chromosome analysis. All other proteins were classified as autosomal. The module cal dn ds from BioPython[67] was used to estimate the non-synonymous and synonymous substitution in the alignment with options ‘method=NG86’.

GO term enrichment analysis

We conducted a GO term enrichment analysis of O. bimaculoides Z genes with the entire genome as the background with DAVID[68]. No functional clusters were found to be significantly enriched.

References

1. Hanlon, R.T., and Messenger, J.B. (2018). Cephalopod Behaviour 2nd ed. (Cambridge University Press) https://doi.org/10.1017/9780511843600.

2. Hanlon, R. (2007). Cephalopod dynamic camouflage. Current Biology 17, R400–R404. https://doi.org/10.1016/j.cub.2007.03.034.

3. Young, J.Z. (1971). The anatomy of the nervous system of Octopus vulgaris (Clarendon Press).

4. Huang, Z., Huang, W., Liu, X., Han, Z., Liu, G., Boamah, G.A., Wang, Y., Yu, F., Gan, Y., Xiao, Q., et al. (2022). Genomic insights into the adaptation and evolution of the nautilus, an ancient but evolving "living fossil." Molecular Ecology Resources 22, 15–27. https://doi.org/10.1111/1755-0998.13439.

5. Ponder, W. ed. (2008). Phylogeny and Evolution of the Mollusca (University of California Press) https://doi.org/10.1525/california/9780520250925.001.0001.

6. Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., and Rokhsar, D.S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524, 220–224. https://doi.org/10.1038/nature14668.

7. Albertin, C.B., Medina-Ruiz, S., Mitros, T., Schmidbaur, H., Sanchez, G., Wang, Z.Y., Grimwood, J., Rosenthal, J.J.C., Ragsdale, C.W., Simakov, O., et al. (2022). Genome and transcriptome mechanisms driving cephalopod evolution. Nat Commun 13, 2427. https://doi.org/10.1038/s41467-022-29748-w.

8. Wang, J., and Zheng, X. (2018). Cytogenetic studies in three octopods, Octopusminor, Amphioctopusfangsiao, and Cistopuschinensis from the coast of China. Comp Cytogenet 12, 373–386. https://doi.org/10.3897/CompCytogen.v12i3.25462.

9. Charlesworth, D., Charlesworth, B., and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. Heredity 95, 118–128. https://doi.org/10.1038/sj.hdy.6800697.

10. Charlesworth, D. (2013). Plant sex chromosome evolution. J Exp Bot 64, 405–420. https://doi.org/10.1093/jxb/ers322.

11. Jonika, M.M., Alfieri, J.M., Sylvester, T., Buhrow, A.R., and Blackmon, H. (2022). Why not Y naught. Heredity 129, 75–78. https://doi.org/10.1038/s41437-022-00543-z.

12. Johnson, N.A., and Lachance, J. (2012). The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. Ann N Y Acad Sci 1256, E1-22. https://doi.org/10.1111/j.1749-6632.2012.06748.x.

13. Perkins, R.D., Gamboa, J.R., Jonika, M.M., Lo, J., Shum, A., Adams, R.H., and Blackmon, H. (2019). A database of amphibian karyotypes. Chromosome Research 27, 313–319. https://doi.org/10.1007/s10577-019-09613-1.

14. Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., et al. (2014). Sex Determination: Why So Many Ways of Doing It? PLOS Biology 12, e1001899. https://doi.org/10.1371/journal.pbio.1001899.

15. The Tree of Sex Consortium (2014). Tree of Sex: A database of sexual systems. Sci Data 1, 140015. https://doi.org/10.1038/sdata.2014.15.

16. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology 20, 238. https://doi.org/10.1186/s13059-019-1832-y.

17. Whitelaw, B.L., Jones, D.B., Guppy, J., Morse, P., Strugnell, J.M., Cooke, I.R., and Zenger, K. (2022). High-Density Genetic Linkage Map of the Southern Blue-ringed Octopus (Octopodidae: Hapalochlaena maculosa). Diversity 14, 1068. https://doi.org/10.3390/d14121068.

18. Kim, B.-M., Kang, S., Ahn, D.-H., Jung, S.-H., Rhee, H., Yoo, J.S., Lee, J.-E., Lee, S., Han, Y.-H., Ryu, K.-B., et al. (2018). The genome of common long-arm octopus Octopus minor. GigaScience 7, giy119. https://doi.org/10.1093/gigascience/giy119.

19. Li, F., Bian, L., Ge, J., Han, F., Liu, Z., Li, X., Liu, Y., Lin, Z., Shi, H., Liu, C., et al. (2020). Chromosome-level genome assembly of the East Asian common octopus (Octopus sinensis) using PacBio sequencing and Hi-C technology. Molecular Ecology Resources 20, 1572–1582. https://doi.org/10.1111/1755-0998.13216.

20. da Fonseca, R.R., Couto, A., Machado, A.M., Brejova, B., Albertin, C.B., Silva, F., Gardner, P., Baril, T., Hayward, A., Campos, A., et al. (2020). A draft genome sequence of the elusive giant squid, Architeuthis dux. GigaScience 9, giz152. https://doi.org/10.1093/gigascience/giz152.

21. Zhang, Y., Mao, F., Mu, H., Huang, M., Bao, Y., Wang, L., Wong, N.-K., Xiao, S., Dai, H., Xiang, Z., et al. (2021). The genome of Nautilus pompilius illuminates eye evolution and biomineralization. Nature Ecology & Evolution, 1–12. https://doi.org/10.1038/s41559-021-01448-6.

22. Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I.T., Novak, A.M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., et al. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature 587, 246–251. https://doi.org/10.1038/s41586-020-2871-y.

23. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15, 1034–1050. https://doi.org/10.1101/gr.3715005.

24. Hobolth, A., Christensen, O.F., Mailund, T., and Schierup, M.H. (2007). Genomic Relationships and Speciation Times of Human, Chimpanzee, and Gorilla Inferred from a Coalescent Hidden Markov Model. PLOS Genetics 3, e7. https://doi.org/10.1371/journal.pgen.0030007.

25. Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., et al. (2012). Insights into hominid evolution from the gorilla genome sequence. Nature 483, 169–175. https://doi.org/10.1038/nature10842.

26. Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K., and Yasunaga, T. (1987). Male-driven Molecular Evolution: A Model and Nucleotide Sequence Analysis. Cold Spring Harb Symp Quant Biol 52, 863–867. https://doi.org/10.1101/SQB.1987.052.01.094.

27. Bergeron, L.A., Besenbacher, S., Zheng, J., Li, P., Bertelsen, M.F., Quintard, B., Hoffman, J.I., Li, Z., St. Leger, J., Shao, C., et al. (2023). Evolution of the germline mutation rate across vertebrates. Nature 615, 285–291. https://doi.org/10.1038/s41586-023-05752-y.

28. Wilson Sayres, M.A., and Makova, K.D. (2011). Genome analyses substantiate male mutation bias in many species. Bioessays 33, 938–945. https://doi.org/10.1002/bies.201100091.

29. Fernández-Álvarez, F.Á., Ibáñez, C., and Díaz-Santana-Iturrios, M. Ibáñez et al. (2021). Macroevolutionary Trade-Offs and Trends in Life History Traits of Cephalopods Through a Comparative Phylogenetic Approach. Frontiers in Marine Science, 8, 1263.

30. Lovell, J.T., Sreedasyam, A., Schranz, M.E., Wilson, M., Carlson, J.W., Harkess, A., Emms, D., Goodstein, D.M., and Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. eLife 11, e78526. https://doi.org/10.7554/eLife.78526.

31. Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40, e49. https://doi.org/10.1093/nar/gkr1293.

32. Wang, Y., Tang, H., Wang, X., Sun, Y., Joseph, P.V., and Paterson, A.H. (2024). Detection of colinear blocks and synteny and evolutionary analyses based on utilization of MCScanX. Nat Protoc 19, 2206–2229. https://doi.org/10.1038/s41596-024-00968-2.

33. Schmidbaur, H., Kawaguchi, A., Clarence, T., Fu, X., Hoang, O.P., Zimmermann, B., Ritschard, E.A., Weissenbacher, A., Foster, J.S., Nyholm, S.V., et al. (2022). Emergence of novel cephalopod gene regulation and expression through large-scale genome reorganization. Nat Commun 13, 2172. https://doi.org/10.1038/s41467-022-29694-7.

34. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics 54, 1.30.1-1.30.33. https://doi.org/10.1002/cpbi.5.

35. Han, W., Liu, L., Wang, J., Wei, H., Li, Y., Zhang, L., Guo, Z., Li, Y., Liu, T., Zeng, Q., et al. (2022). Ancient homomorphy of molluscan sex chromosomes sustained by reversible sex-biased genes and sex determiner translocation. Nat Ecol Evol 6, 1891–1906. https://doi.org/10.1038/s41559-022-01898-6.

36. Evensen, K.G., Robinson, W.E., Krick, K., Murray, H.M., and Poynton, H.C. (2022). Comparative phylotranscriptomics reveals putative sex differentiating genes across eight diverse bivalve species. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 41, 100952. https://doi.org/10.1016/j.cbd.2021.100952.

37. Bonnaud, L., Ozouf-Costaz, C., and Boucher-Rodoni, R. (2004). A molecular and karyological approach to the taxonomy of Nautilus. Comptes Rendus Biologies 327, 133–138. https://doi.org/10.1016/j.crvi.2003.12.004.

38. Vitturi, R., Rasotto, M.B., and Farinella-Ferruzza, N. (1982). The chromosomes of 16 molluscan species. Italian Journal of Zoology 49, 61–71.

39. Gao, Y.M., and Natsukari, Y. (1990). Karyological Studies on Seven Cephalopods. Venus (Japanese Journal of Malacology) 49, 126–145. https://doi.org/10.18941/venusjjm.49.2_126.

40. Iwasaki, M., Kajiwara, T., Yasui, Y., Yoshitake, Y., Miyazaki, M., Kawamura, S., Suetsugu, N., Nishihama, R., Yamaoka, S., Wanke, D., et al. (2021). Identification of the sex-determining factor in the liverwort Marchantia polymorpha reveals unique evolution of sex chromosomes in a haploid system. Curr Biol 31, 5522-5532.e7. https://doi.org/10.1016/j.cub.2021.10.023.

41. Carey, S.B., Jenkins, J., Lovell, J.T., Maumus, F., Sreedasyam, A., Payton, A.C., Shu, S., Tiley, G.P., Fernandez-Pozo, N., Healey, A., et al. (2021). Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. Sci Adv 7, eabh2488. https://doi.org/10.1126/sciadv.abh2488.

42. Toups, M.A., and Vicoso, B. (2023). The X chromosome of insects likely predates the origin of class Insecta. Evolution 77, 2504–2511. https://doi.org/10.1093/evolut/qpad169.

43. Songco-Casey, J.O., Coffing, G.C., Piscopo, D.M., Pungor, J.R., Kern, A.D., Miller, A.C., and Niell, C.M. (2022). Cell types and molecular architecture of the Octopus bimaculoides visual system. Curr Biol 32, 5031-5044.e4. https://doi.org/10.1016/j.cub.2022.10.015.

44. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770. https://doi.org/10.1093/bioinformatics/btr011.

45. Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun 11, 1432. https://doi.org/10.1038/s41467-020-14998-3.

46. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods 18, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

47. Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics 36, 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025.

48. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst 3, 95–98. https://doi.org/10.1016/j.cels.2016.07.002.

49. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., et al. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356, 92–95. https://doi.org/10.1126/science.aal3327.

50. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA 117, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

51. Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics Chapter 4, Unit 4.10. https://doi.org/10.1002/0471250953.bi0410s05.

52. Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics and Bioinformatics 3, lqaa108. https://doi.org/10.1093/nargab/lqaa108.

53. Tardaguila, M., de la Fuente, L., Marti, C., Pereira, C., Pardo-Palacios, F.J., del Risco, H., Ferrell, M., Mellado, M., Macchietto, M., Verheggen, K., et al. (2018). SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res 28, 396–411. https://doi.org/10.1101/gr.222976.117.

54. Kuo, R.I., Cheng, Y., Zhang, R., Brown, J.W.S., Smith, J., Archibald, A.L., and Burt, D.W. (2020). Illuminating the dark side of the human transcriptome with long read transcript sequencing. BMC Genomics 21, 751. https://doi.org/10.1186/s12864-020-07123-7.

55. Dainat, J. (2022). Another Gtf/Gff Analysis Toolkit (AGAT): Resolve interoperability issues and accomplish more with your annotations. Plant and Animal Genome XXIX Conference. https://github.com/NBISweden/AGAT.

56. Hubisz, M.J., Pollard, K.S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. Brief Bioinform 12, 41–51. https://doi.org/10.1093/bib/bbq072.

57. Vasimuddin, Md., Misra, S., Li, H., and Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 314–324. https://doi.org/10.1109/IPDPS.2019.00041.

58. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

59. Pedersen, B.S., and Quinlan, A.R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 34, 867–868. https://doi.org/10.1093/bioinformatics/btx699.

60. Song, W., Li, R., Zhao, Y., Migaud, H., Wang, C., and Bekaert, M. (2021). Pharaoh Cuttlefish, Sepia pharaonis, Genome Reveals Unique Reflectin Camouflage Gene Set. Frontiers in Marine Science 8.

61. Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations. Bioinformatics 37, 1639–1643. https://doi.org/10.1093/bioinformatics/btaa1016.

62. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

63. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 28, 1747–1756. https://doi.org/10.1101/gr.239244.118.

64. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

65. Ranwez, V., Douzery, E.J.P., Cambon, C., Chantret, N., and Delsuc, F. (2018). MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. Molecular Biology and Evolution 35, 2582–2584. https://doi.org/10.1093/molbev/msy159.

66. Ranwez, V., Chantret, N., and Delsuc, F. (2021). Aligning Protein-Coding Nucleotide Sequences with MACSE. Methods Mol Biol 2231, 51–70. https://doi.org/10.1007/978-1-0716-1036-7_4.

67. Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25, 1422–1423. https://doi.org/10.1093/bioinformatics/btp163.

68. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44–57. https://doi.org/10.1038/nprot.2008.211.

Bridge

In Chapter III, I documented my discovery of the *O. bimaculoides* Z chromosome facilitated by a de-novo chromosome-level genome assembly. Using a comparative genomics approach we showed that the Z chromosome is shared across all extant cephalopod species, making it one of the oldest known sex determination systems.

In Chapter VI, I pivot my questions from the macro to the micro-evolutionary scale by exploring the population genetics of two closely related octopus species, *O. bimaculoides* and *O. bimaculatus.* Like my first two chapters, I continue to leverage the new genome assembly that I generated by using it as a reference assembly for calling variants in both species. We found that the population genetics statistics and estimates of demographic history are consistent with the life history and reproductive strategy differences between the two species.

CHAPTER IV: DIVERSITY AND DIVERGENCE OF TWO SYMPATRIC, SIBLING OCTOPUS SPECIES

Co-authors of this chapter include Dr. Silas Tittes, Dr. Scott T. Small, and Dr. Andrew D. Kern. All authors contributed to experimental questions and design. A.D.K oversaw the project. G.C.C. generated and filtered VCFs, calculated population genetics statistics, and estimated demographic history. S.T. estimated recombination rates. S.T., S.T.S, and A.D.K advised on data curation and filtering. G.C.C. prepared the initial draft of this manuscript, all authors contributed to the writing and editing of the final manuscript.

Introduction

The commonly named Two-spot octopus consists of two sympatric, sister species: *Octopus bimaculoides* Pickford & McConnaughey (1949)[1] and *Octopus bimaculatus* Verrill (1883)[2]. These species are characterized by blue eye spots, or ocelli, located beneath the eyes near the base of arms two and three[1,3]. Due to their nearly identical external appearances, it was not until 1949 that the Two-spot octopus was recognized as two distinct species[1]. However, despite their similar physical traits, these sister species are differentiated by their geographic ranges and reproductive strategies (Table 1). It also has been suggested that there are premating barriers to reproduction between species as male *O. bimaculoides* will not copulate with female *O. bimaculatus* and vice versa[1]. Generally, the species can be differentiated by the overall shape of the ocelli and the numbers of suckers per arm. In *O. bimaculatus*, the ocelli contain blue spokes like that of a wheel and have longer arms with 140 – 190 suckers per arm[4]. In *O. bimaculoides* the ocelli contain a ring of blue dots that resemble chain links and their arms have

200 – 300 suckers[5,6]. The hectocotylized arm is relatively shorter in *O. bimaculatus* and *O. bimaculoides* has a longer sperm reservoir and shorter spermatophore[7]. At sexual maturity, *O. bimaculatus* is generally larger than *O. bimaculoides*[1].

Both Two-spot octopus species live in rocky reefs, muddy areas, and sandy bottom habitats. The range of *O. bimaculoides* is more northerly, extending from the coast of southern California, U.S.A. to Magdalena Bay on the Baja California peninsula, Mexico[3,8,9]. In contrast, the range of *O. bimaculatus* begins in southern California but extends further south into the northern half of the Gulf of California[5,8,10]. It has been suggested that the two species may occupy slightly different ecological niches as *O. bimaculoides* is frequently found in the high intertidal *Laminaria* zone, where seaweed beds begin[1]. These zones contain rocky tidal flats, coastal lagoons, shallow bays, and mudflat areas that are often exposed at low tide[6]. *O. bimaculatus*, on the other hand, is typically found in deeper parts of the intertidal and subtidal communities, below the *Laminaria* zone[1]. *O. bimaculatus* are often caught and studied near California's Channel Islands[4,11,12]. *O. bimaculoides* has been found as deep as 20 meters and *O. bimaculatus* has been found as deep as 50 meters[4,6].

The sister species also have differing reproductive seasons, fecundity, and egg characteristics. *O. bimaculoides* lay relatively small festoons, or chain-like structures, of large eggs that develop directly into juvenile, benthic hatchling octopuses. In contrast, *O. bimaculatus* lay clutches of over 20,000 small eggs that have a planktonic paralarval drift stage that allows individuals to disperse far from their hatching sites over a period of up to three months[8,13]. In both species the timing of the reproductive season depends on the geographic location. Generally, in shared geographic ranges their reproductive periods are non-overlapping[8] (Table 1).

The first research study that identified the Two-spot octopus as two different species found that they are infected by different parasites[1]. The study found that *O. bimaculoides* is infected with *Dicyemennea californica* and *O. bimaculatus* is infected by *Dicyemenna abelis.* Both parasites are a species of dicymenid that inhabits cephalopod renal sacs. There is no record of dicymenidae causing harm to their cephalopod hosts, but it is possible that they are beneficial[14].

As marine stocks are continuously fished at unsustainable levels around the world, it is predicted that humans will become more reliant on resources in lower trophic levels, such as cephalopods[15,16]. The total share of global cephalopod trade has been increasing despite inadequate management practices. In 2022, cephalopods made up 7 percent of the world's total exports of aquatic animal products[17]. There is no commercial fishing of the Two-spot octopus in the U.S., however many small-scale fishers of Northwest Mexico sell octopus catch locally and in commercial markets[8]. The difficulty of classifying octopus catch at the species level prevents official reports from including this information in their fishery statistics[8]. Without species-level data, it impossible to know whether a particular population is being overexploited. Due to the different life history strategies in the sister Two-spot octopus species, it is possible that they require distinct management strategies.

Our current understanding of cephalopod evolution has mostly focused on long time scales concerning large taxonomic groups[18–20], with emphasis on the traits that make the cephalopod lineage unique from their invertebrate relatives (i.e. large nervous systems, camouflage abilities, unique body plan), while less is known about microevolutionary processes for any particular species (but see[8,21–24]). One study examined genetic diversity and population structure using 16S DNA and COI sequences of *O. bimaculoides* and *O. bimaculatus* individuals

from Northwest Mexico. They found that *O. bimaculatus* has higher levels of gene flow and a higher effective population size than *O. bimaculoides*[8]. We aimed to expand on these findings by investigating population structure, diversity and divergence, demographic history, and evidence of selection in Two-spot octopus individuals from the California coast. To our knowledge, this is the first population genetics study of a cephalopod using whole genomic sequencing data.

Life history	***O. bimaculoides***	***O. bimaculatus***
Geographic distribution	From San Simeon, CA, USA to Bahia San Quintin in BC, Mexico	From CA, USA to Bahia Vizcaino BCS, including Gulf of California
Reproductive period	Santa Barbara, CA, USA (Dec-May) San Quintin, BCP, Mexico (Oct-Jan)	Pacific coast of BCP (Jan-Jun) Gulf of California (Jun-Sep)
Fecundity	Festoons of 137 – 780 eggs	Clutches of > 20,000 eggs
Egg size	10-12mm	4-7mm
Planktonic larval duration	Absent, direct development to juvenile, benthic hatchlings	2-3 months
Size at sexual maturity	55 mm (ML) males 110 mm (ML) females	125 mm (ML) males 147 mm (ML) females
Lifespan	1.0-1.5 years	1.5-2.0 years

Table 4.1: Summary of life history strategies of the Two-spot octopus species. This table was adapted from Dominguez-Contreras et al. 2018.

Materials and Methods

Sample collection and resequencing

O. bimaculoides and *O. bimaculatus* samples were collected from the southern coast of California between 2019 and 2022. Various tissue types were either flash frozen or stored in RNAlater (Table S4.1). DNA extraction, sample preparation, and sequencing for all samples was conducted by the University of Oregon Genomics & Cell Characterization Core Facility (GC3F). A separate sequencing library was prepared for each species. Genomic DNA was extracted,

pooled, and sequenced on two Illumina NovaSeq 6000 S4 lanes. In total, our study includes six *O. bimaculoides* and thirteen *O. bimaculatus* individuals (Table S4.1). Sequencing data from four of the *O. bimaculatus* individuals was published previously[25].

Variant calling

We used SNPArcher[26] to call variants separately in each species using the newest *O. bimaculoides* genome assembly[25] as the reference. SNPArcher maps reads to the reference with bwa-mem[27] and generates a VCF using GATK[28]. We conducted our own filtering scheme on SNPArcher's raw output VCFs using BCFtools[29]. We filtered out all repetitive sites with the repeat map that was generated with the Repeat Masker[30] and Repeat Modeler[31] pipeline in Coffing et al. 2025. We also set genotypes to missing (i.e. ./.) according to depth or the allele depth to depth ratio, depending on the genotype. Next, we removed sites that did not pass a set of thresholds; quality below 20, quality by depth below 10, mapping quality below 40, genotype missingness above 20%, depth higher than 2x the overall average depth. We additionally removed all indels. All sites that did not pass our filtering scheme were included in a mask bed file that was used in downstream analyses. Overall, filtering resulted in 1.77 million SNPs in the *O. bimaculoides* VCF and 14.59 million SNPs in the *O. bimaculatus* VCF. In addition to the individual VCFs and masks for each species, we generated a merged VCF with 14.52 million SNPs using BCFtools merge and a merged mask using BEDtools merge[32]. In all final VCFs, we kept invariant sites.

Depth calculations

Using the bam files that were generated with SNPArcher[26], we calculated read depth in 500,000 bp windows using samtools depth[33]. We examined depth across chromosomes for each

individual and found that two of the *O. bimaculatus* individuals that were marked as female (Oblat_F_7 and Oblat_F_8) did not have half coverage in the Z chromosome. This led us to conclude there was an error in sexing these individuals in the field and we adjusted their sex label to Male (Table S4.1).

Calculation of population genetics statistics

We calculated nucleotide diversity (π), Tajima's D, d_{XY}, and Weir and Cockerham F_{ST} in non-overlapping 1Mb windows using the python package scikit-allel[34]. The mask we generated from our filtering scheme (see above) was used as input for the negative accessibility mask. Population statistics were calculated on the independent VCFs for each species. We used scikit-allel to generate PCA plots for each chromosome from the merged VCF using 40,000 random biallelic SNPs with a minor allele count of at least two.

Demographic analysis

We used SMC++ (v.1.15.2)[35] to estimate the history of effective population size and divergence time of the two octopus species. We input the merged VCF, containing 14.52 million SNPs, into the vcf2smc command using the merged mask as input for missing data. This output a separate smc file for each chromosome. The estimate command of SMC++ was used to estimate population size history of each species with a mutation rate of 2.4×10^{-9} as reported for the Southern blue-ringed octopus (*Hapalochlaena maculosa*)[36], which is the closest related species with a reported mutation rate estimate. To infer divergence time of the two species we independently generated a joint site frequency spectrum for each species using the vcf2smc command. Finally, the command split was run to refine the marginal estimates to infer the joint demography. Demographic history was estimated independently for each of the 30 octopus

chromosomes. The median and upper and lower quantiles were estimated from the 30 indpendent history traces.

Estimating recombination landscapes and the effect of linked selection

We used ReLERNN[37] (https://github.com/kr-colab/relernn_nextflow) to infer local recombination rates along the genome separately for *O. bimaculoides* and *O. bimaculatus.* For efficiency, training and predictions were split up into 12Mb windows along the chromosomes and we provided the same accessibility mask as used in estimating the population genetic summary statistics to account for missing data (see above). We tested a difference in genome-wide recombination rate between the two species using a Mann-Whitney U test. Estimating recombination allowed us to assess how local nucleotide diversity varies with recombination rate among the two species, which in turn allowed us to assess the relative impacts of linked selection acting on the genome of the two species. We did this by combining nucleotide diversity in 1Mb windows with estimates of local recombination rate, along with the density of coding sequences in the same window.

Results

Population structure

We generated whole genome sequences from a total of 19 octopus individuals, including 6 *O. bimaculoides* and 13 *O. bimaculatus*, from the coast of southern California. Principal component analysis (PCA) supports a clear separation of the two species (Figure 4.1). On chromosome 1, PC1 and PC2 account for 70.8% and 5.7% of total variation between the two species, respectively. PCA plots for all other chromosomes show similar patterns in variance to

chromosome 1 (Figures S4.1 – S4.3). The variance in PC2 shows a separation of two *O. bimaculatus* individuals, Oblat_F_8 and Oblat_F_5, from the remaining samples.

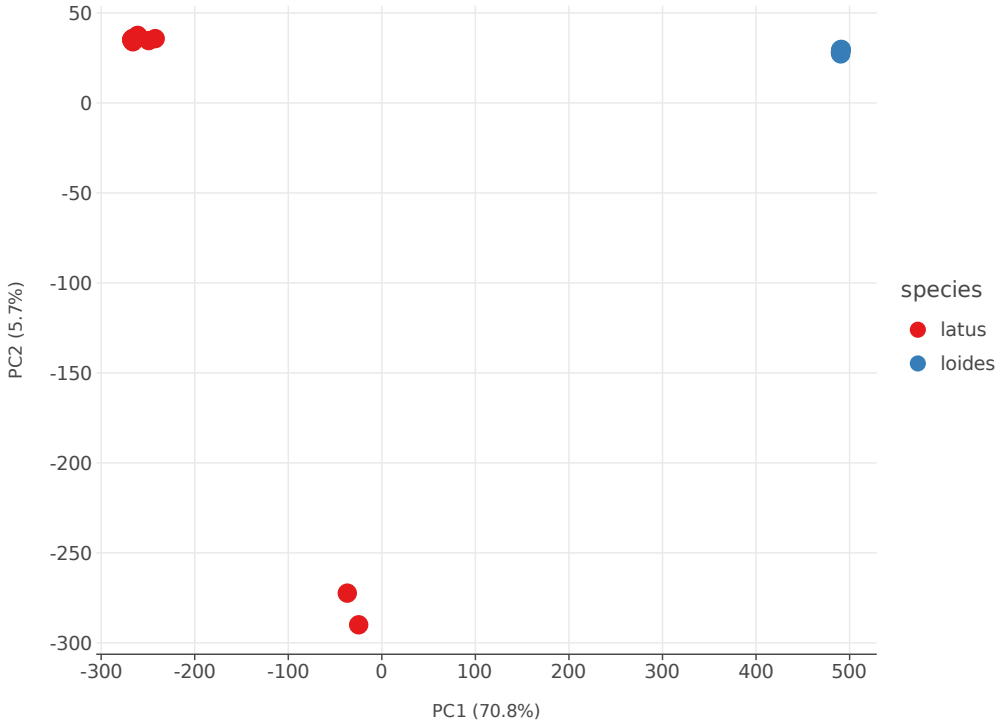

Figure 4.1: Principal component analysis (PCA) of chromosome 1 from *O. bimaculoides* and *O. bimaculatus* samples.

Within-species diversity and between-species divergence

We find that across the genome, nucleotide diversity (π) is higher in *O. bimaculatus* than in *O. bimaculoides* (Mann-Whitney U test, $p < 0.001$, Figures 4.2 & S4.4 – S4.62). Average nucleotide diversity is 7.12×10^{-4} and 2.601×10^{-3} in *O. bimaculoides* and *O. bimaculatus* respectively. We also find that the distribution of Tajima's D is significantly different between the two species (Mann-Whitney U test, $p = 0.0$), with *O. bimaculoides* showing a slightly positive average (0.489) and *O. bimaculatus* a slightly negative average (- 0.189). We estimated divergence across chromosomes with d_{XY} and find the average to be 0.012, indicating few fixed

differences between the two species. F_{ST} appears to be high (average windowed = 0.705) between the two species, but we note this is likely inflated because of the low nucleotide diversity in *O. bimaculoides*[38,39]. On the Z chromosome, we observed lower nucleotide diversity in both species compared to the autosomes (Mann-Whitney U test, $p < 0.001$). We also find a lower d_{XY} (Mann-Whitney U test, $p < 0.001$) and a higher F_{ST} compared to the autosomes (Mann-Whitney U test, $p < 0.001$, Figures 4.2 & S4.63 – S4.72). This mirrors the divergence estimates we reported in Coffing et al. (2025), where we show that divergence on the Z chromosome was depressed relative to autosomes in a deeper timescale comparison between *O. bimaculoides* and *O. sinensis*.

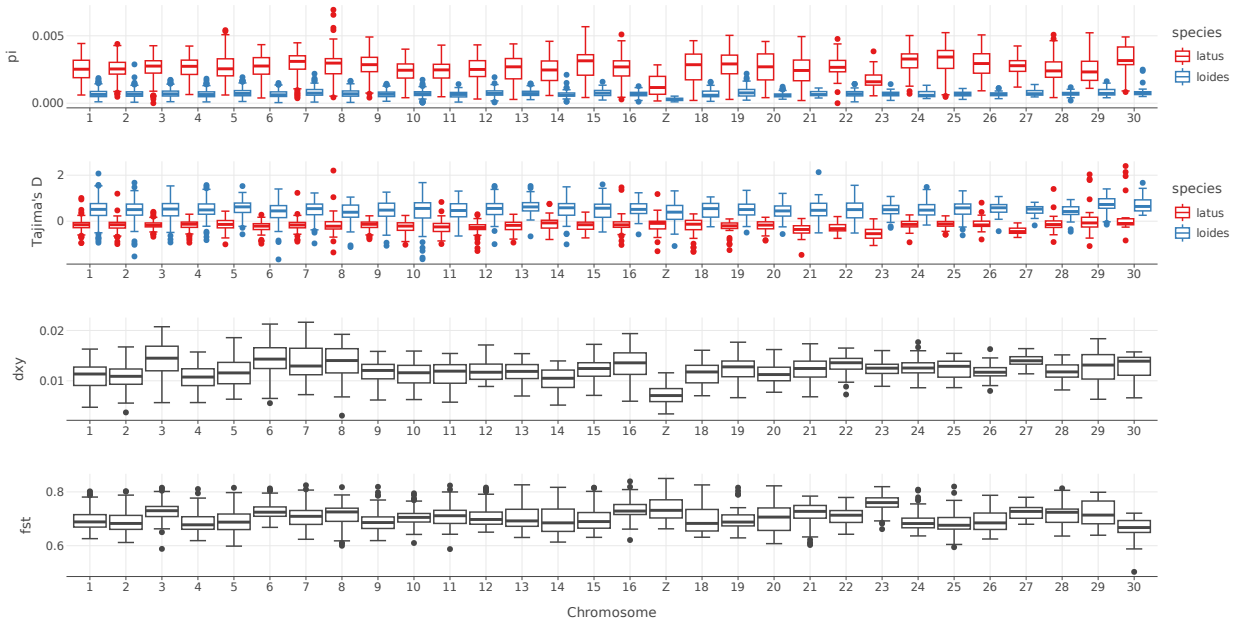

Figure 4.2: Population genomics statistics calculated for *O. bimaculoides* (loides) and *O. bimaculatus* (latus).

<u>Octopus demographic history</u>

To investigate the demographic history and estimate split time between our octopus species, we used SMC++[35]. The results show a history of a larger effective population size in *O. bimaculatus* than *O. bimaculoides* (Fig. 4.3). SMC++ infers population splits by implementing a "clean split" model in which it is assumed that there is no gene flow after the split of the populations. The estimated split time is as long as around 400,000 generations (800,000 years) ago (Fig. 4.3). The uncertainty of this estimation is represented by the overlap of the red and blue confidence intervals between 250,000 and 400,000 generations ago. Overall, we see that the effective population size of *O. bimaculoides* was in slight decline after the split, a population size increase around 63,000 generations ago, and then another decline. In contrast, the population size of *O. bimaculatus* has been relatively stable. SMC++ histories are in good agreement with our summary of the site frequency spectrum (Tajima's D; above) where the slightly positive average *D* found in *O. bimaculoides* reflects the history of population decline, and values near zero for *O. bimaculatus* are indicative of population size stasis.

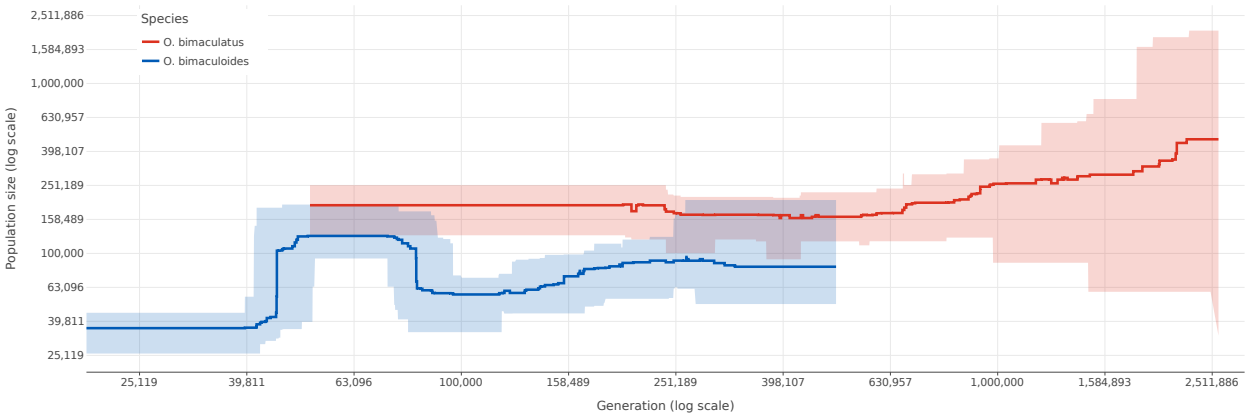

Figure 4.3: Demographic history estimations generated with SMC++ for *O. bimaculoides* and *O. bimaculatus*. The dark blue and dark red lines are the median demographic history estimations from all 30 chromosomes. Upper and lower quartiles estimated from the 30 chromosomes are represented by the light blue and light red ribbons.

Recombination rate and patterns of linked selection

We next set out to use our chromosome-scale population genetic data to estimate rates of recombination using the deep learning estimator ReLERNN[37]. Estimates reveal that *O. bimaculoides* and *O. bimaculatus* have very similar rates of recombination, 3.92×10^{-9} and 3.78×10^{-9}, respectively, although the overall window mean estimates are significantly higher in *O. bimaculoides* (Mann-Whitney U test, p=0.0032; Figure 4.4A). At the per chromosome level correlations between rates of recombination and levels of polymorphism have long been used to reveal the action of linked selection in genomes[40,41]. Among the Two-spot octopus species we see a striking difference – while *O. bimaculatus* shows a stronger positive correlation between polymorphism and recombination, *O. bimaculoides* does not (Figure 4.4B). Such a correlation is expected under many models of linked selection, wherein higher rates of recombination allow neutral variation to decouple from selected variation nearby, ultimately preserving variation linked to positively or negatively selected mutations[41–43]. Despite differences in the relationship between diversity and recombination, the relationship between nucleotide diversity and the density of coding sequences within the same windows is quite similar between the two species (Figure 4C), however we reemphasize that *O. bimaculoides* has globally less variation, so the dynamic range of the comparison is not equivalent between species. We provide further interpretation for these results in the Discussion.

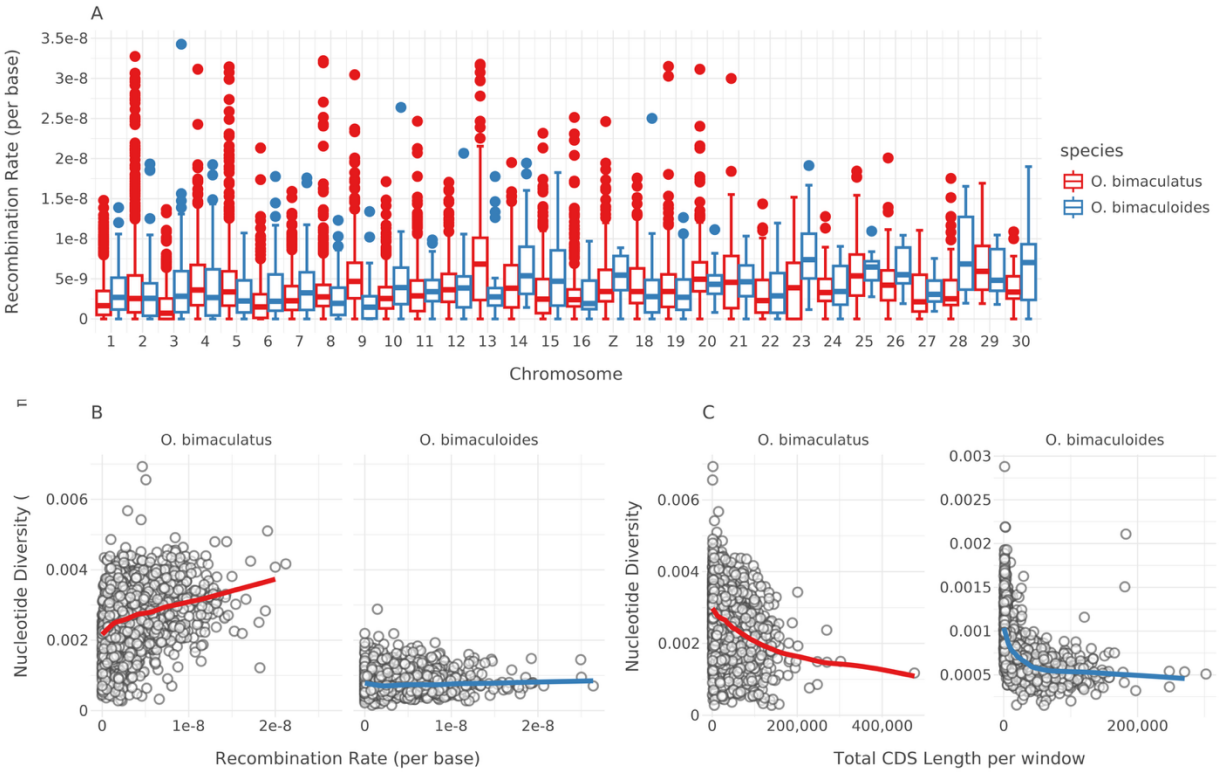

Figure 2.4: Recombination and linked selection for *O. bimaculoides* and *O. bimaculatus*.
(A) Recombination rate variation across chromosomes for *O. bimaculoides* (blue) and *O. bimaculatus* (red).
(B) The relationship between nucleotide diversity (π) and recombination rate for *O. bimaculoides* (blue) and *O. bimaculatus* (red).
(C) The relationship between nucleotide diversity (π) and density of coding sequences for *O. bimaculoides* (blue) and *O. bimaculatus* (red). Loess trend lines were added to both panels B and C.

Discussion

<u>Estimates of diversity, divergence, and demographic history consistent with life history differences between Two-spot octopus species</u>

We found *O. bimaculoides* to have lower nucleotide diversity (π) and a consistently lower effective population size than *O. bimaculatus* (Figures 4.2 and 4.3). These estimates are consistent with the differences in life history and reproductive strategies between these species as well as the patterns of previous findings that analyzed mitochondrial genes and microsatellite loci[8]. *O. bimaculatus* has a planktonic paralarval drift stage, unlike *O. bimaculoides,* which

provides more opportunity for dispersal from the hatching site and can potentially increase gene flow between populations. Additionally, *O. bimaculatus* lays far more eggs per clutch than *O. bimaculoides*, which could contribute to a larger population size and higher nucleotide diversity levels. Tajima's D was found to be slightly positive in *O. bimaculoides,* which is consistent with a history of a declining effective population size, whereas the near zero average Tajima's D in *O. bimaculatus* is consistent with our estimated constant-sized population history.

Genetic differentiation between the species is relatively low and indicative of a recent split time, as expected. For instance our estimates of d_{XY} between the Two-spot octopus species (average windowed d_{XY} = 0.012) is roughly on the order of human-chimpanzee divergence (approximately 0.01)[44]. Assuming the mutation rate of μ=2.4e-9 per bp per generation, as estimated from *H. maculosa*, our average d_{XY} estimate implies divergence roughly 2.5 million generations ago (~ 5M years). This is likely an overestimate of the split time, as population histories derived using an SMC estimator (SMC++) show estimates that are quite a bit younger (from ~500k-800k years).

We can also compare absolute measures of divergence with relative measures. In comparison to d_{XY}, our estimates of F_{ST} across the genome are quite high (average windowed = 0.705, Figure 4.2). F_{ST} is a measure of the proportion of genetic diversity that is attributable to differences among populations. However, F_{ST} is well known to be sensitive to levels of within-population variation (e.g., nucleotide diversity (π)[45]. As *O. bimaculoides* has globally reduced levels of nucleotide diversity compared to *O. bimaculatus,* we see quite high values of F_{ST} in comparison to d_{XY}. Overall, what emerges is a history of relatively recent divergence among sister clades, one of whom (*O. bimaculoides)* suffered a significant population decline after speciation.

There is currently no evidence of hybridization between the Two-spot octopus species. A preliminary study placed a female *O. bimaculoides* in a tank with a male *O. bimaculatus* individual (and vice versa) and found that they neither interacted or made any mating attempts, suggesting tentative evidence of pre mating isolation[1]. Although this study was rudimentary and contained a small sample size, its results combined with a lack of genomic evidence for hybridization, including high levels of F_{ST}, suggest behavioral barriers to mating may exist. These results raise questions about the nature of pre- and post-zygotic barriers and the factors that might be driving speciation of these two species. Disentangling the nature and timing of speciation between *O. bimaculoides* and *O. bimaculatus,* especially the mechanisms of reproductive isolation and the role of natural selection and geographic separation, is a valuable area for further research.

Z chromosome polymorphism

Sex chromosomes are expected to show distinct evolutionary dynamics compared to autosomes, largely due to differences in selection pressures, genetic drift, and effective population sizes[46,47]. The Fast-Z effect (or Fast-X effect in XY systems) predicts that coding regions of the Z chromosome will evolve faster than those on autosomes, leading to higher F_{ST} – a relative measure of genetic differentiation between populations – on the Z compared to autosomes, a pattern found in several species[47–51]. One theory theorized to drive this pattern is that the heterogametic sex exposes recessive mutations to selection – intensifying both purifying selection against deleterious alleles and positive selection on beneficial alleles[47].

Between the Two-spot octopus species, we found F_{ST} to be higher on the Z chromosome compared to autosomes, which is consistent with the Fast-Z effect hypothesis. However, in

contrast to this pattern, we observed a significantly lower d_{XY} and reduced nucleotide diversity on the Z chromosomes compared to the autosomes in both species. The combination of these findings is consistent with the hypothesis that Z chromosome is prone to purifying selection of recessive, mildly deleterious alleles, which quickly removes diversity[47–49]. These results are also consistent with our previous findings that divergence is reduced on the Z chromosomes when looking at deeper time scales between *O. bimaculoides* and *O. sinensis* (Chapter III, Figure 3.2B). Additionally, the lower population size of the Z chromosome may contribute to these patterns as well. Overall, our patterns are similar to those found in closely-related greenish warbler species in which the Z chromosomes have a lower nucleotide diversity and d_{XY} compared to autosomes combined with a higher F_{ST} [52]. We have not identified any regions of the Z chromosome to be under faster positive selection than the autosomes using McDonald-Kreitman tests, but dN/dS ratios and sweep scans still need to be analyzed.

Stronger drift likely limits selection in *O. bimaculoides*

We found recombination rates to be higher in *O. bimaculoides* than in *O. bimaculatus.* Estimating recombination rates allowed us to search for signatures of linked selection, where neutral diversity reveals the impact of nearby selected alleles[38,39,41–43]. We see patterns consistent with linked selection in *O. bimaculatus* but not *O. bimaculoides* (Fig 4.4). The differences between species are likely explained by a lower population size in *O. bimaculoides,* which limits the efficiency of natural selection to modify the allele frequencies of non-neutral mutations.

Conclusions

To our knowledge, we have presented the first whole genome re-sequencing dataset of a cephalopod in a pair of sympatric, sibling octopus species. Notably, these species possess an ancient ZZ/ZO sex determination system, offering a novel framework to test the theory of the

Fast-Z effect. Future studies might be able to identify sex-linked traits or explore in depth how sex chromosome evolution might contribute to reproductive isolation and speciation in these octopuses. Finally, as fishing of cephalopods is increasing, it will be critical to monitor the population sizes and diversity of these species independently.

References

1. Pickford, G., and McConnaughey, B. (1949). The *Octopus bimaculatus* problem: a study in sibling species. Bulletin of the Bingham Oceanographic Collection 12.

2. Verrill, A.E. (1883). Descriptions of two species of Octopus from California. Museum of Comparative Zoology.

3. Morris, R.H., Abbott, D.P., and Haderlie, E.C. (1980). Intertidal invertebrates of California (Stanford University Press).

4. Hofmeister, J.K.K., and Voss, K.M. (2024). *Octopus bimaculatus*, California two-spot octopus. In Octopus Biology and Ecology (Elsevier), pp. 31–45. https://doi.org/10.1016/B978-0-12-820639-3.00012-1.

5. Lang, M.A., and Hochberg, F.G. (1997). *Octopus bimaculoides.* In M. A. Lang, & F. G. Hochberg (Eds.), The fishery and market potential of octopus in California (proceedings of a workshop at the University of Southern California Catalina Marine Science Center, 31 Aug. to 8 Sept., 1989). In (Washington, DC: Smithsonian Institution), pp. 1–9.

6. Markaida, U., and Castellanos-Martínez, S. (2024). *Octopus bimaculoides*, Lesser two-spotted octopus. In Octopus Biology and Ecology (Elsevier), pp. 47–59. https://doi.org/10.1016/B978-0-12-820639-3.00017-0.

7. Voight, J.R. (2001). The relationship between sperm reservoir and spermatophore length in benthic octopuses (Cephalopoda: Octopodidae). Journal of the Marine Biological Association of the United Kingdom 81, 983–986. https://doi.org/10.1017/S0025315401004945.

8. Domínguez-Contreras, J.F., Munguia-Vega, A., Ceballos-Vázquez, B.P., Arellano-Martínez, M., García-Rodríguez, F.J., Culver, M., and Reyes-Bonilla, H. (2018). Life histories predict genetic diversity and population structure within three species of octopus targeted by small-scale fisheries in Northwest Mexico. PeerJ 6, e4295. https://doi.org/10.7717/peerj.4295.

9. Alejo-Plata, M. del C., Salgado-Ugarte, I., Herrera-Galindo, J., and Meraz-Hernando, J. (2014). Cephalopod biodiversity at Gulf of Tehuantepec, Mexico, determinate from direct sampling and diet analysis on large pelagic-fishes predators. Hidrobiologica 24, 57–68.

10. Ambrose, R.F. (1997). *Octopus bimaculatus.* In R.F. Ambrose (Eds.), The fishery and market potential of octopus in California (proceedings of a workshop at the University of Southern California Catalina Marine Science Center, 31 Aug. to 8 Sept., 1989). In (Washington, DC: Smithsonian Institution), pp. 11–22.

11. Ambrose, R. (1988). Population dynamics of *Octopus bimaculatus:* influence of life history patterns, synchronous reproduction and recruitment. Malacologia 29, 23–39.

12. Hofmeister, J.K.K., and Voss, K.M. (2017). Activity space and movement patterns of *Octopus bimaculatus* (Verrill, 1883) around Santa Catalina Island, California. Journal of Experimental Marine Biology and Ecology 486, 344–351. https://doi.org/10.1016/j.jembe.2016.10.024.

13. Villanueva, R., Vidal, E.A.G., Fernández-Álvarez, F.Á., and Nabhitabhata, J. (2016). Early Mode of Life and Hatchling Size in Cephalopod Molluscs: Influence on the Species Distributional Ranges. PLOS ONE 11, e0165334. https://doi.org/10.1371/journal.pone.0165334.

14. Furuya, H., and Souidenne, D. (2019). Dicyemids. In Handbook of Pathogens and Diseases in Cephalopods, C. Gestal, S. Pascual, Á. Guerra, G. Fiorito, and J. M. Vieites, eds. (Springer International Publishing), pp. 159–168. https://doi.org/10.1007/978-3-030-11330-8_11.

15. Pauly, D., Christensen, V., Guénette, S., Pitcher, T.J., Sumaila, U.R., Walters, C.J., Watson, R., and Zeller, D. (2002). Towards sustainability in world fisheries. Nature 418, 689–695. https://doi.org/10.1038/nature01017.

16. Sala, E., Aburto-Oropeza, O., Reza, M., Paredes, G., and López-Lemus, L.G. (2004). Fishing Down Coastal Food Webs in the Gulf of California. Fisheries 29, 19–25. https://doi.org/10.1577/1548-8446(2004)29[19:FDCFWI]2.0.CO;2.

17. The State of World Fisheries and Aquaculture 2024 (2024). (FAO) https://doi.org/10.4060/cd0683en.

18. Albertin, C.B., Medina-Ruiz, S., Mitros, T., Schmidbaur, H., Sanchez, G., Wang, Z.Y., Grimwood, J., Rosenthal, J.J.C., Ragsdale, C.W., Simakov, O., et al. (2022). Genome and transcriptome mechanisms driving cephalopod evolution. Nat Commun 13, 2427. https://doi.org/10.1038/s41467-022-29748-w.

19. Chen, Z., Baeza, J.A., Chen, C., Gonzalez, M.T., González, V.L., Greve, C., Kocot, K.M., Arbizu, P.M., Moles, J., Schell, T., et al. (2025). A genome-based phylogeny for Mollusca is concordant with fossils and morphology. Science 387, 1001–1007. https://doi.org/10.1126/science.ads0215.

20. Schmidbaur, H., Kawaguchi, A., Clarence, T., Fu, X., Hoang, O.P., Zimmermann, B., Ritschard, E.A., Weissenbacher, A., Foster, J.S., Nyholm, S.V., et al. (2022). Emergence of novel cephalopod gene regulation and expression through large-scale genome reorganization. Nat Commun 13, 2172. https://doi.org/10.1038/s41467-022-29694-7.

21. Bein, B., Lima, F.D., Lazzarotto, H., Rocha, L.A., Leite, T.S., Lima, S.M.Q., and Pereira, R.J. (2023). Population genomics of an Octopus species identify oceanographic barriers and inbreeding patterns. Mar Biol 170, 161. https://doi.org/10.1007/s00227-023-04307-z.

22. Lau, S.C.Y., Wilson, N.G., Watts, P.C., Silva, C.N.S., Cooke, I.R., Allcock, A.L., Mark, F.C., Linse, K., Jernfors, T., and Strugnell, J.M. (2025). Circumpolar and Regional Seascape Drivers of Genomic Variation in a Southern Ocean Octopus. Molecular Ecology 34, e17601. https://doi.org/10.1111/mec.17601.

23. McKeown, N.J., Arkhipkin, A.I., and Shaw, P.W. (2019). Genetic analysis reveals historical and contemporary population dynamics in the longfin squid *Doryteuthis gahi*: implications for cephalopod management and conservation. ICES Journal of Marine Science 76, 1019–1027. https://doi.org/10.1093/icesjms/fsz009.

24. Cheng, S.H., Gold, M., Rodriguez, N., and Barber, P.H. (2021). Genome-wide SNPs reveal complex fine scale population structure in the California market squid fishery (Doryteuthis opalescens). Conserv Genet 22, 97–110. https://doi.org/10.1007/s10592-020-01321-2.

25. Coffing, G.C., Tittes, S., Small, S.T., Songco-Casey, J.O., Piscopo, D.M., Pungor, J.R., Miller, A.C., Niell, C.M., and Kern, A.D. (2025). Cephalopod sex determination and its ancient evolutionary origin. Current Biology 35, 931-939.e4. https://doi.org/10.1016/j.cub.2025.01.005.

26. Mirchandani, C.D., Shultz, A.J., Thomas, G.W.C., Smith, S.J., Baylis, M., Arnold, B., Corbett-Detig, R., Enbody, E., and Sackton, T.B. (2024). A Fast, Reproducible, High-throughput Variant Calling Workflow for Population Genomics. Molecular Biology and Evolution 41, msad270. https://doi.org/10.1093/molbev/msad270.

27. Vasimuddin, Md., Misra, S., Li, H., and Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 314–324. https://doi.org/10.1109/IPDPS.2019.00041.

28. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110.

29. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience 10, giab008. https://doi.org/10.1093/gigascience/giab008.

30. Smit, A.F.A., Hubley, R., and Green, P. (2013). RepeatMasker Open-4.0. http://www.repeatmasker.org.

31. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences 117, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

32. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

33. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

34. Miles, A., Bot, P.I., Rodrigues, M.F., Ralph, P., Harding, N., Pisupati, R., and Rae, S. (2024). scikit-allel. Version v1.3.13.

35. Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole-genomes. Nat Genet 49, 303–309. https://doi.org/10.1038/ng.3748.

36. Whitelaw, B.L., Jones, D.B., Guppy, J., Morse, P., Strugnell, J.M., Cooke, I.R., and Zenger, K. (2022). High-Density Genetic Linkage Map of the Southern Blue-ringed Octopus (Octopodidae: Hapalochlaena maculosa). Diversity 14, 1068. https://doi.org/10.3390/d14121068.

37. Adrion, J.R., Galloway, J.G., and Kern, A.D. (2020). Predicting the Landscape of Recombination Using Deep Learning. Molecular Biology and Evolution 37, 1790–1808. https://doi.org/10.1093/molbev/msaa038.

38. Hahn, M.W. (2019). Molecular population genetics (Sinauer associates).

39. Cruickshank, T.E., and Hahn, M.W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology 23, 3133–3157. https://doi.org/10.1111/mec.12796.

40. Begun, D.J., and Aquadro, C.F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356, 519–520. https://doi.org/10.1038/356519a0.

41. Corbett-Detig, R.B., Hartl, D.L., and Sackton, T.B. (2015). Natural Selection Constrains Neutral Diversity across A Wide Range of Species. PLOS Biology 13, e1002112. https://doi.org/10.1371/journal.pbio.1002112.

42. Buffalo, V. (2021). Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's Paradox. eLife 10, e67509. https://doi.org/10.7554/eLife.67509.

43. Lewontin, R.C. (1974). The genetic basis of evolutionary change (Columbia Univ. Pr).

44. Rodrigues, M.F., Kern, A.D., and Ralph, P.L. (2024). Shared evolutionary processes shape landscapes of genomic variation in the great apes. Genetics 226, iyae006. https://doi.org/10.1093/genetics/iyae006.

45. Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution 15, 538–543. https://doi.org/10.1093/oxfordjournals.molbev.a025953.

46. Vicoso, B., and Charlesworth, B. (2006). Evolution on the X chromosome: unusual patterns and processes. Nat Rev Genet 7, 645–653. https://doi.org/10.1038/nrg1914.

47. Charlesworth, B., Coyne, J.A., and Barton, N.H. (1987). The Relative Rates of Evolution of Sex Chromosomes and Autosomes. The American Naturalist 130, 113–146. https://doi.org/10.1086/284701.

48. Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., et al. (2012). The genomic landscape of species divergence in Ficedula flycatchers. Nature 491, 756–760. https://doi.org/10.1038/nature11584.

49. Oyler-McCance, S.J., Cornman, R.S., Jones, K.L., and Fike, J.A. (2015). Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. Heredity 115, 452–459. https://doi.org/10.1038/hdy.2015.46.

50. Mank, J.E., Axelsson, E., and Ellegren, H. (2007). Fast-X on the Z: Rapid evolution of sex-linked genes in birds. Genome Res. 17, 618–624. https://doi.org/10.1101/gr.6031907.

51. Wright, A.E., Harrison, P.W., Zimmer, F., Montgomery, S.H., Pointer, M.A., and Mank, J.E. (2015). Variation in promiscuity and sexual selection drives avian rate of Faster-Z evolution. Molecular Ecology 24, 1218–1235. https://doi.org/10.1111/mec.13113

52. Irwin, D.E., Alcaide, M., Delmore, K.E., Irwin, J.H., and Owens, G.L. (2016). Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. Molecular Ecology 25, 4488–4507. https://doi.org/10.1111/mec.13792.

CHAPTER V: CONCLUSION

Cephalopods harbor an extraordinary collection of morphological novelties including camera-type eyes, complex nervous systems, unique body plans, and camouflage abilities. This combination of unique traits has made cephalopods a foundational study system in diverse areas of biology from neuroscience to ecology. With the increased accessibility to genome sequencing, there has been an explosion in the field of cephalopod genomics in recent years revealing that cephalopods have evolutionary novelties within their genomes as well as in their morphology. My dissertation adds to our growing knowledge of cephalopod biology by examining the evolutionary history of the octopus visual system, octopus sex determination, and octopus speciation.

In Chapter II, I presented our work to determine the transcriptional cell types of the octopus visual system. Through convergent evolution, octopuses and vertebrates have evolved anatomically similar eyes despite 500 million years since divergence. However, the octopus brain is organized differently than vertebrates and little was known about the molecular determinants of the octopus visual system. To determine the cell types, we performed single-cell RNA sequencing on the *Octopus bimaculoides* optic lobe, the primary center for visual processing in the brain. Accurate analysis of the single-cell data required that generate a high quality, chromosome-level genome assembly of the octopus. The new assembly enabled us to generate a single cell RNA-seq atlas revealing six major neuronal classes identified by neurotransmitter expression and non-neuronal and immature neuron cell populations. Overall, this work provides a map to be used in the investigation of the role of distinct cell types in visual processing.

In Chapter III, I discovered that the *O. bimaculoides* possess a sex chromosome, which was the first evidence of genetic sex determination in a cephalopod. By using the high-quality genome assembly that I generated in Chapter II, along with additional sequencing data, I demonstrated that the female octopus has just one sex chromosome, consistent with a ZO karyotype, while the male has two (ZZ). I examined the repeat landscape of the Z chromosome, finding that it has a significantly higher LINE density compared to the autosomes. Through careful genome comparisons, I show that this LINE density pattern is found in single chromosomes in other cephalopod species, including another octopus species, a squid, and a cuttlefish. Synteny analyses showed that these chromosomes with elevated LINE densities are orthologous across cephalopods, which lead me to hypothesize that they were sex chromosomes. I analyzed publically available coverage data from a representative species in each cephalopod clade (i.e. octopus, squid, cuttlefish, and nautilus) and found that the putative sex chromosomes were at half coverage in each species. Together, this evidence suggested that the Z chromosomes is shared across all extant cephalopod species, making it over ~480 million years old.

In Chapter IV, I examined the microevolutionary patterns of *O. bimaculoides* and its sister, sympatric species *O. bimaculatus.* I generated sequencing data from several individuals in both species and called variant sites. With this data, I asked whether diversity, divergence, and demographic history of these sister species is consistent with their life history strategies. I calculated population genetics statistics including nucleotide diversity (π), Tajima's D, d_{XY}, and F_{ST}. I found that diversity is lower in *O. bimaculoides* and that divergence levels show few fixed differences between the species. Demographic history estimates shows that since the split time, the population size of *O. bimaculoides* has been in decline. Finally, recombination rates are

similar in both species, but the overall window mean estimates are higher in *O. bimaculoides*. Overall, my findings show that these octopus species have distinct evolutionary patterns.

In this dissertation, I addressed a diverse set of evolutionary questions using multiple types of sequencing datasets. I addressed the limitations of analyzing a single-cell RNA sequencing dataset in Chapter II by generating a chromosome-level octopus genome assembly. As a result of this assembly, I discovered that cephalopods have a ZZ/ZO sex determination system in Chapter III. In Chapter IV, I used the genome assembly as a tool to analyze the microevolutionary patterns of two sister octopus species. Together, these projects illuminate both the deep evolutionary history and more recent divergence of an emerging model organism. Broadly, my work demonstrates the power of integrative genomic approaches to uncover fundamental biological insights from a non-traditional study system. I hope that the findings of my research will inspire future investigations into the unique evolutionary trajectories of cephalopods.

APPENDIX A
SUPPORTING INFORMATION FOR CHAPTER II

Methods S2.1. Elucidating unidentified genes. Related to Figures 2.5, 2.5, and 2.7.

In a number of cases, genes of interest were not well annotated using orthogroup assignment or simple BLAST searches to the non-redundant database. Here we describe additional steps we took to try to assign likely functions to our candidates. In most of these cases orthologous protein predictions could be retrieved from other cephalopod genomes, but not beyond, suggesting that these uncharacterized genes are novel and potentially specific to cephalopods or octopuses. With these in hand, we sought to predict function via deep homology searching as well as structure and function prediction. Homology search and function prediction included the use of the hidden Markov model based tools HHblits and HHpreds[S2.1,S2.2], and PSI-BLAST[S2.3]. In each case we predicted protein structures using ColabFold[S4], which itself is based on AlphaFold2.0[S2.5]. With these predicted structures in hand, we then searched for distant structural homologues using foldseek[S2.6].

obimac0010569

Obimac0010569 aligns well to an uncharacterized protein found in O. vulgaris and O. sinensis called LOC115219258. This hit however combines two separate predicted genes from the previous O. bimaculoides genome: obimac0010569 and obimac0010570. While we have Iso-Seq reads that cover obimac0010570, we have none that link obimac0010569 and obimac0010570. Assuming our gene model has split this locus in error, we created multiple species alignments of the candidate and used the resulting multiple species alignment (MSA) for homology search. HHblits revealed a strong signal of a conserved m13 domain on c-terminal

side of the predicted protein, with aligns to proteins like putative m13 family peptidases (e-vals ~ 1e-100) and endothelin converting enzymes (e-vals ~ 1e-90). HHPred showed similar signals, with peptidases and in particular metallopeptidases showing homology on the c-terminal end. Interestingly there is some weak homology to Protocadherin-15, but e-values were all >> 0.01

obimac0022194

We were able to find obimac0022194 orthologs from *O. sinensis, O. vulgaris,* and *O. minor.* Using a multiple sequence alignment from these orthologs as input to HHblits revealed hits to uncharacterized proteins from the cuttlefish, *S. pharaonis* (e-val ~ 2.2e-15), the limpet *Lottia gigantea* (e-val ~ 3.9e-8), and slightly weaker hits to a protein annotated as luquin1 in the land slug *Deroceras reticulatum* (e-val 9.9e-8) and a protein annotated as luquin neuropeptide fragment in the annelid polychaete *Platynereis dumerilii* (e-val 2.6e-5). HHpred gave no further insight. We then used AlphaFold2.0 to computationally predict a structure using our MSA. The resulting structures were generally low confidence and largely disordered. A representative view of the rank1 model is shown below in Figure ST2.1. We further searched for structural homologs using foldseek, but zero hits were returned.

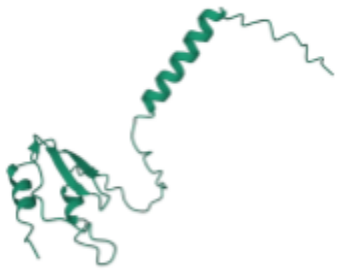

Figure ST2.1. AlphaFold prediction for obimac0022194 showing the structure of the rank1 model.

obimac0019980

Using protein BLAST searches against individual cephalopod genomes we were able to find an O. sinensis ortholog to obimac0019980 with strong homology (e-val ~ 3e-88) and only a very weak hit to the cuttlefish *Sepia officinalis* (e-val = 0.003). We used pairwise alignments of the *O. bimaculoides* and *O. sinensis* proteins as input to HHblits and HHpred. HHblits recovered no significant homology. HHpred similarly did not recover orthologs in distant genomes but did predict a transmembrane domain and a signal peptide sequence within obimac0019980. We used DeepTMHMM[S2.7] to confirm the presence of the signal peptide and transmembrane domain. A clear signal peptide is present and DeepTMHMM classifies this protein as globular + signal peptide (Figure ST2.2). We further used AlphaFold2.0 to provide a predicted structure. This structure was largely disordered but two clear helical regions are revealed. Topology searching using foldseek revealed no similar structures.

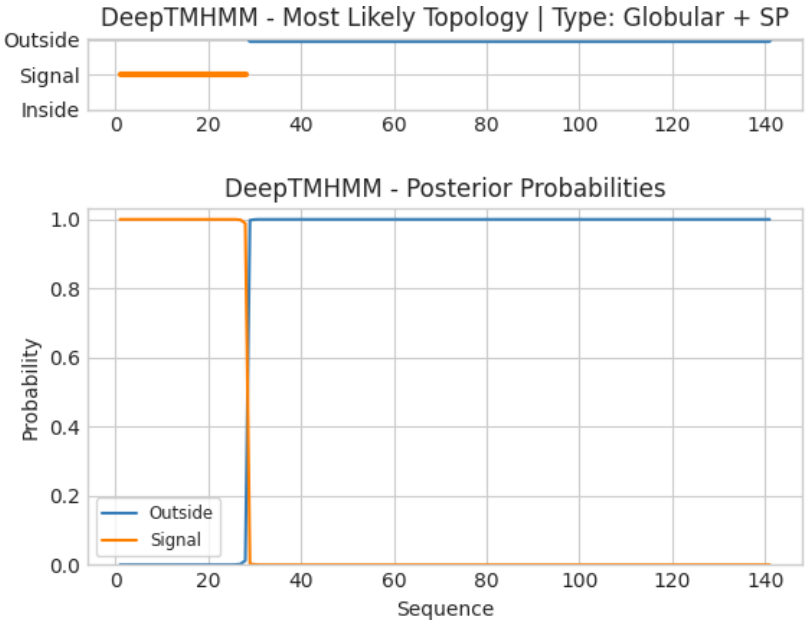

Figure ST2.2. Evidence of a globular protein plus signal peptide structure for obimac0019980 as predicted by DeepTMHMM[S2.7].

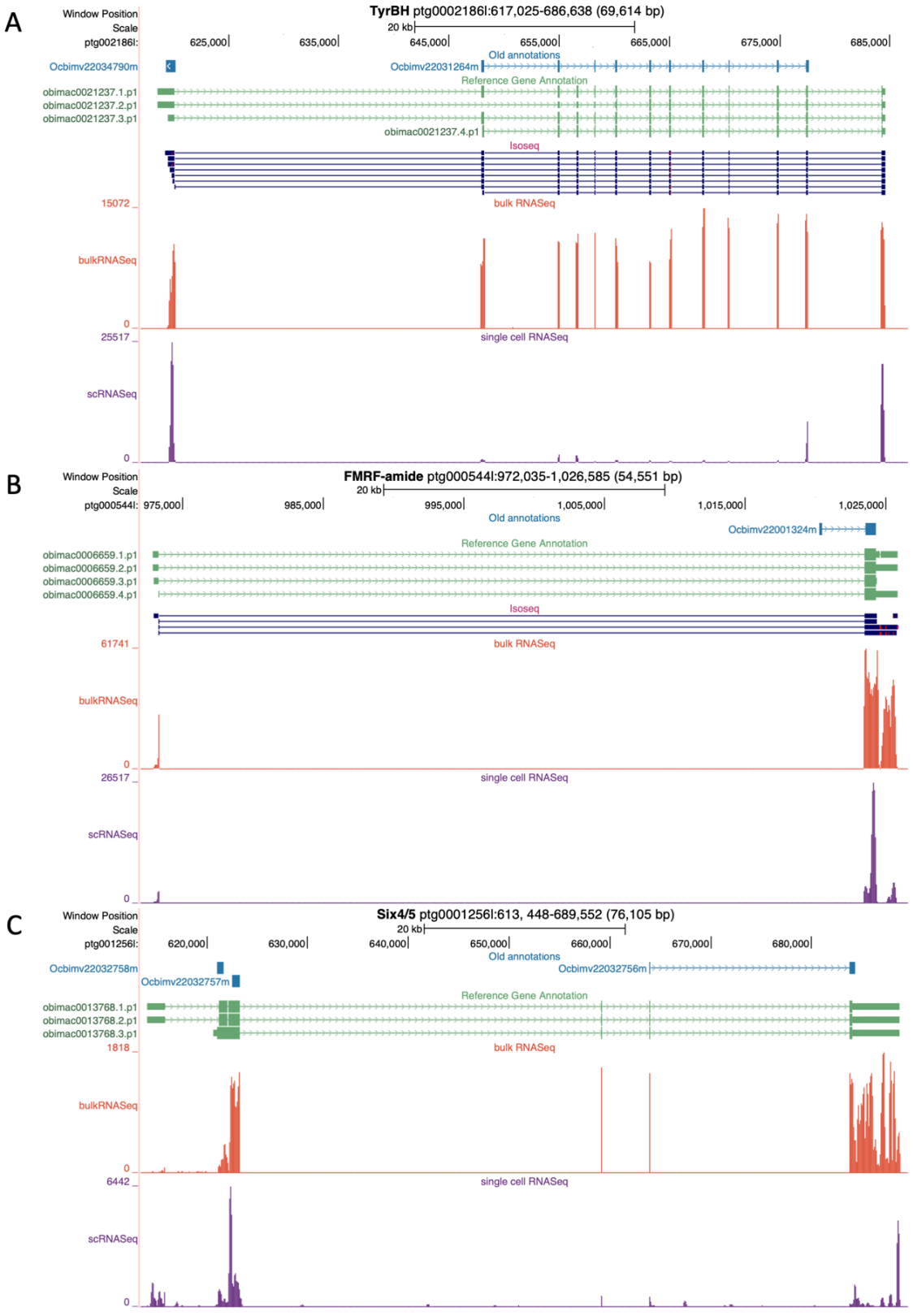
A
Window Position
Scale
ptg002186l:
TyrBH ptg0002186l:617,025-686,638 (69,614 bp)
20 kb
625,000
635,000
645,000
655,000
665,000
675,000
685,000
Old annotations
Ocbimv22034790m
Ocbimv22031264m
Reference Gene Annotation
obimac0021237.1.p1
obimac0021237.2.p1
obimac0021237.3.p1
obimac0021237.4.p1
Isoseq
15072
bulk RNASeq
bulkRNASeq
0
25517
single cell RNASeq
scRNASeq
0
B
Window Position
Scale
ptg000544l:
FMRF-amide ptg000544l:972,035-1,026,585 (54,551 bp)
20 kb
975,000
985,000
995,000
1,005,000
1,015,000
1,025,000
Old annotations
Ocbimv22001324m
Reference Gene Annotation
obimac0006659.1.p1
obimac0006659.2.p1
obimac0006659.3.p1
obimac0006659.4.p1
Isoseq
61741
bulk RNASeq
bulkRNASeq
0
26517
single cell RNASeq
scRNASeq
0
C
Window Position
Scale
ptg001256l:
Six4/5 ptg0001256l:613, 448-689,552 (76,105 bp)
20 kb
620,000
630,000
640,000
650,000
660,000
670,000
680,000
Old annotations
Ocbimv22032758m
Ocbimv22032757m
Ocbimv22032756m
Reference Gene Annotation
obimac0013768.1.p1
obimac0013768.2.p1
obimac0013768.3.p1
1818
bulk RNASeq
bulkRNASeq
0
6442
single cell RNASeq
scRNASeq
0

Figure S2.1. Genome browser output showing improved gene models. Related to Figure 2.2. Three genes are shown to demonstrate examples of the substantial improvement our gene models achieved with an updated annotation. Each panel contains the following tracks: old gene annotations (blue) from the original *O. bimaculoides* genome annotation all isoforms of new annotations (green), Iso-seq alignment data (dark blue, if present), bulk RNA sequencing alignment data (orange), and single-cell RNA sequencing alignment data (purple). Our single-cell data is poly-A captured, meaning most reads align to the 3' ends of genes or areas of the genome.
(A) ***tyrbh*** New gene models (labeled in green, with each potential isoform represented by the obimac gene identifier followed by a period and numeric) include two additional exons compared to the original annotation, one on each end of the gene. These new exons are supported by both Iso-Seq data and bulk RNA-seq data. By including both of these exons, we have captured two piles of scRNA-seq read alignments on each end of the gene.
(B) ***flri.*** New models lengthened the 3' ends and included an additional exon on the 5' end compared to the original model. Both of these changes to this gene are supported by Iso-Seq bulk RNA-seq read alignments and improved the amount of scRNA-seq data we are able to capture.
(C) ***six4/5***. This gene was not present in our Iso-Seq data, but shows an example of where the new gene model substantially improved annotation of this region. The new models lengthened the 3' end and added two exons in the 5' direction, allowing us to capture more of the scRNA-seq data. We were able to stitch together three genes from the old model (*Ocbimv22032756m, Ocbimv22032757m,* and *Ocbimv22032758m*) into a single gene model (*obimac0013768*). If we had used the old gene model for our single cell analysis, then it would appear that three separate genes have similar expression patterns.

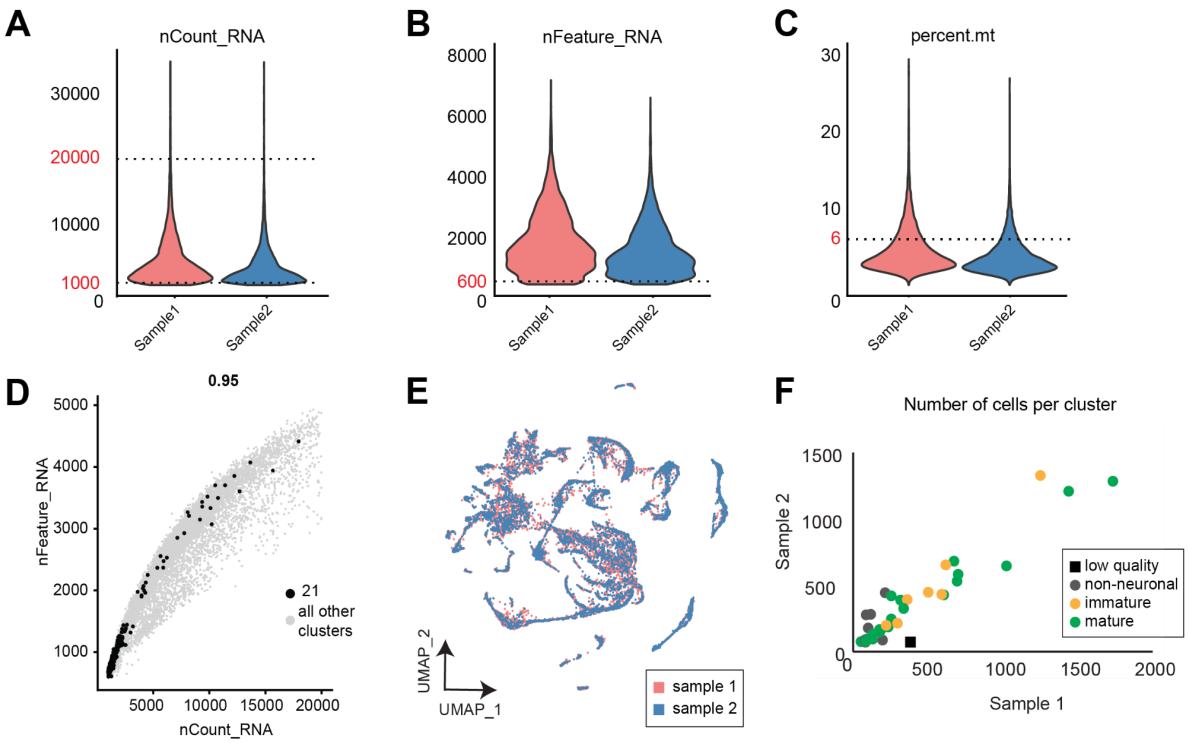

Figure S2.2. Single-cell RNA sequencing quality control metrics. Related to Figure 2.2.
(A-C) Violin plot showing raw reads (nCount_RNA), genes (nFeature_RNA), and mitochondrial percentage (percent.mt) across the two biological replicates used in this study. Dashed lines and red text indicate thresholds set for downstream processing.
(D) Scatter plot showing the correlation between the number of reads and the number of genes for all cells. One cluster (21) had low reads and gene counts and did not have a specific gene expression profile (data not shown). Cluster 21 was therefore excluded from further analysis.
(E) UMAP showing representation of each sample across clusters after applying sctransform normalization and integration via Canonical Correlation Analysis.
(F) Scatter plot of number of cells per cluster, with each biological replicate plotted on each axis. Points are color-coded based on whether they represent clusters from non-neuronal, immature, or mature neuronal populations. Neuronal cell types were represented similarly in the two replicates. On the other hand non-neuronal clusters were more variable, perhaps due to their presence in neighboring tissue. The "low quality" cluster (21), which was discarded from analysis, is represented by a black square and was only present in one replicate.

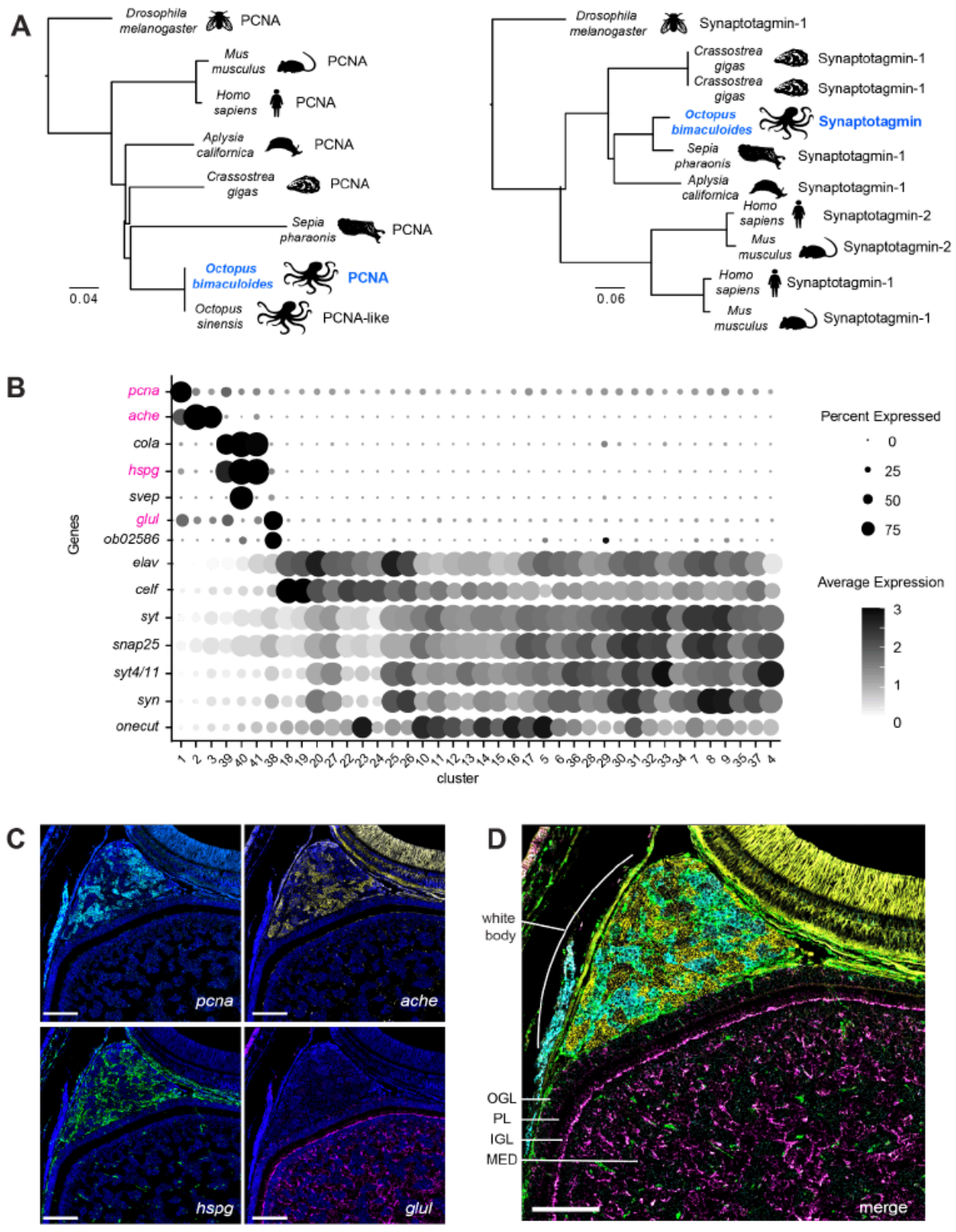
A
Drosophila melanogaster
PCNA
Mus musculus
PCNA
Homo sapiens
PCNA
Aplysia californica
PCNA
Crassostrea gigas
PCNA
Sepia pharaonis
PCNA
Octopus bimaculoides
PCNA
Octopus sinensis
PCNA-like
0.04
Drosophila melanogaster
Synaptotagmin-1
Crassostrea gigas
Synaptotagmin-1
Crassostrea gigas
Synaptotagmin-1
Octopus bimaculoides
Synaptotagmin
Sepia pharaonis
Synaptotagmin-1
Aplysia californica
Synaptotagmin-1
Homo sapiens
Synaptotagmin-2
Mus musculus
Synaptotagmin-2
Homo sapiens
Synaptotagmin-1
Mus musculus
Synaptotagmin-1
0.06
B
Genes
pcna
ache
cola
hspg
svep
glul
ob02586
elav
celf
syt
snap25
syt4/11
syn
onecut
cluster
Percent Expressed
0
25
50
75
Average Expression
3
2
1
0
C
pcna
ache
hspg
glul
D
white body
OGL
PL
IGL
MED
merge

Figure S2.3. Characterization of non-neuronal clusters. Related to Figure 2.2.
(A) Phylogenetic trees from OrthoFinder, demonstrating the method for assignment of gene identities in our updated gene model.
(B) Dot plot of conserved markers, showing demarcation of neuronal vs. non-neuronal cells, as well as subtypes within non-neuronal cells. One subset (clusters 1-3) of non-neuronal cells has relatively high expression of markers relating to proliferation and blood (proliferating cell nuclear antigen (*pcna*)[S2.8] and acetylcholinesterase (*ache*)[S2.9]), whereas a second subset (clusters 39-41) are marked by relatively high expression of genes relating to endothelium (collagen type 1 alpha 1 (cola)[S2.10], heparan sulfate proteoglycan 2 (*hspg*), and sushi von willebrand factor type a (*svep*)). Finally, a third cluster (38) has high expression of glutamine synthase (*glul*)[S.2.11], which is expressed in glia. Interestingly, this cluster also has an uncharacterized gene *obimac0002586*. Markers for these clusters are contrasted with neuronal markers/clusters below.
(C) FISH of genes to demonstrate anatomical locations of non-neuronal cell types. In mapping putative non-neuronal cell clusters to their anatomical locations via FISH, we found several markers to be primarily expressed outside of the optic lobe, mostly in the white body, which is involved in hematopoiesis[S2.12]. *pcna* (top left) marks dividing cells and is localized to the white body, along with ache (top right), which is a marker of blood cells. Together, *pcna* and *ache* expression is consistent with the suggested role of the white body in hematopoiesis. *hspg* (bottom left) demarcates endothelial cells and is prominent in the white body as well as the optic lobe, where their morphology resembles vasculature. One prominent marker gene from these clusters that is expressed in the optic lobe is *glul* (bottom right), a glial marker, which identified a population of cell bodies at the boundary of the plexiform layer and IGL, with expression extending into the neuropils of the medulla, consistent with glial organization in other invertebrate nervous systems[S2.13]. *glul* can be seen within the optic lobe, prominently labeling nuclei on the superficial border of IGL but also extending throughout the neuropil of the medulla. Scale bars indicate 100 um. Nuclei are stained in DAPI.
(D) A quadruple FISH of the genes shown in C without nuclei staining. OGL, outer granular layer; PL, plexiform layer; IGL, inner granular layer; MED, medulla. Scale bar indicates 100 um.

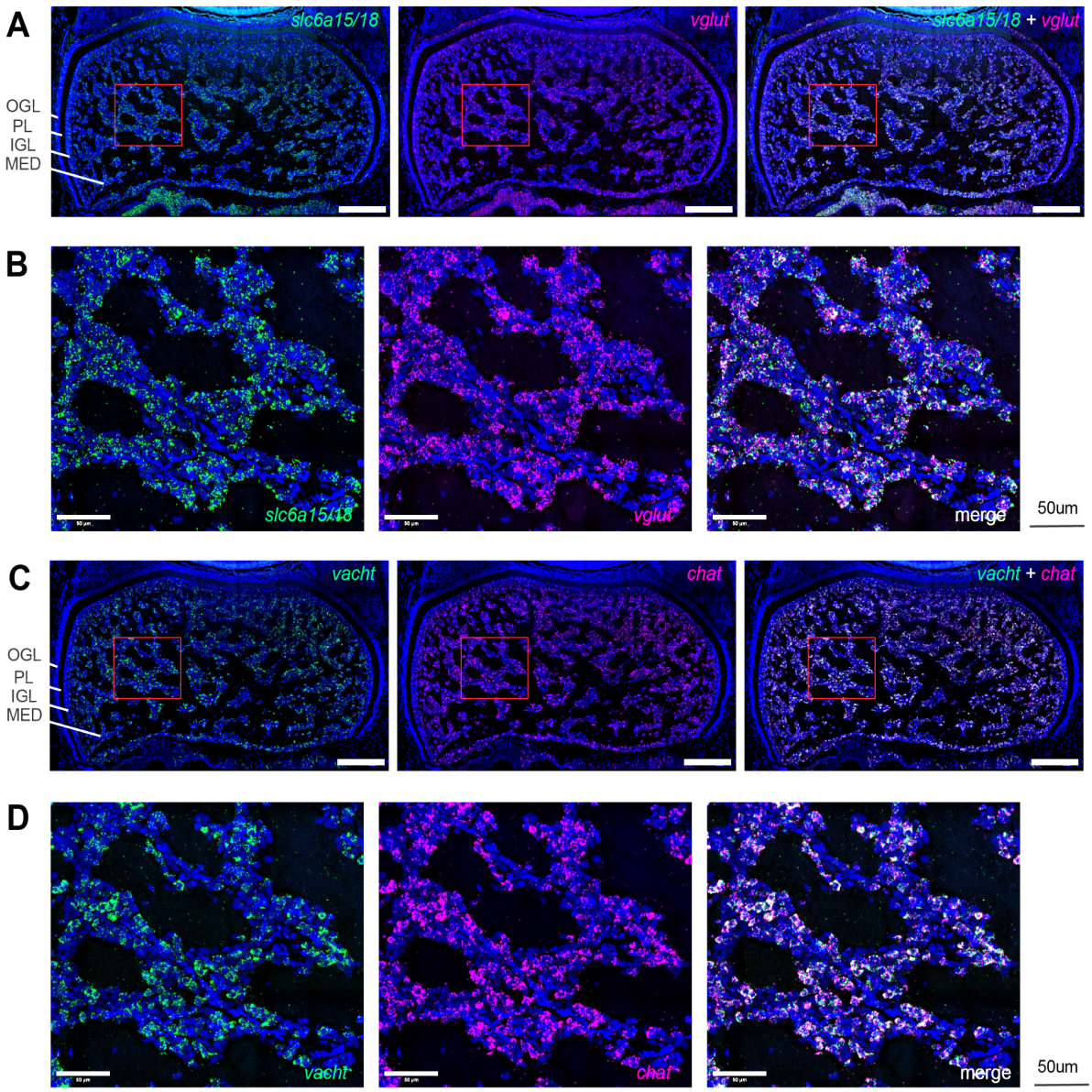

Figure S2.4. Validation of glutamatergic and cholinergic markers. Related to Figures 2-3.
(A-B) 2-color FISH of two glutamatergic markers, *slc6a15/18* and *vglut*, demonstrating co-expression in the optic lobe. Red box in (A) denotes a zoomed-in portion of the medulla in (B). Here and below, nuclei are stained with DAPI and the scale bar represents 200um unless otherwise noted.
(C-D) 2-color FISH of two cholinergic markers, *vacht* and *chat*, demonstrating co-expression. Red box in (C) denotes the region depicted in (D).

Assembly statistic	Octopus_bimaculoides_v2_0 (2015 assembly)	o_bimaculoides_hifi_v1.0.0 (2022 assembly)
Size	2.34 Gb	2.34 Gb
Number of contigs	151,674	5,437
Average contig length	15.42 Kb	430.95 Kb
Longest contig	4.06 Mb	7.63 Mb
Contig N50	475 Kb (n = 1333)	880.44 Kb (n = 755)
Number of gaps	548,450	5 (115 Ns)
Coverage	~60x	34x
BUSCO (Eukaryota)	C:91.8%[S:90.6%,D:1.2%]F:6.7 %,M:1.5%,n:255	C:92.2%[S:90.6%,D:1.6%],F:6.7 %,M:1.1%,n:255
Number of annotated genes	33,638	18,896

Table S2.1. Genome assembly statistics for o_bimaculoides_hifi_v1.0.0 and Octopus_bimaculoides_v2_0. Related to Figure 2.2. The computational pipeline that led to these results is described in Methods. Benchmarking Universal Single-Copy Orthologs (BUSCO v. 3)[S2.14] was run against the eukaryota_odb9 database to show an overall completeness of 92.2% (C: complete, S: complete and single-copy, D: complete and duplicated, F: fragmented, M: missing). We used the Eukaryote database rather than Mollusc or Metazoa because those databases contain the original O. bimaculoides proteome, thus biasing results. The older gene model contains many fragmented gene annotations, which led to an inflated number of total genes in the original gene annotation file. The new genome annotation contains fewer genes because many of the fragmented annotations have been connected.

Gene	Obimac ID	Gene	Obimac ID
ache	obimac0020279	*ob19980*	obimac0019980
bib	obimac0026684	*ob22194*	obimac0022194
cacng	obimac0027453	*octr*	obimac0000086
celf	obimac0028033	*onecut*	obimac0002891
chat	obimac0030402	*orc*	obimac0002637
cola	obimac0002036	*otx*	obimac0030353
dat	obimac0021555	*pcdh-ob09200*	obimac0009200
dchs	obimac0020142	*pcdh-ob26462*	obimac0026462
dlx	obimac0024037	*pcna*	obimac0019754
dscam	obimac0017416	*plxna*	obimac0017384
efn	obimac0011807	*pxfv*	obimac0025590
elav	obimac0020958	*sema2*	obimac0002144
epha	obimac0017206	*sema5*	obimac0023141
flri	obimac0006659	*six4/5*	obimac0013768
fmrf1	obimac0015407	*slc6a15/18*	obimac0030399
gad	obimac0007601	*snap25*	obimac0010589
glul	obimac0032988	*sox2*	obimac0011874
hh	obimac0025144	*svep*	obimac0000862
hspg	obimac0022794	*syn*	obimac0011778
lrrc15	obimac0011126	*syt*	obimac0014441
lxgkr	obimac0011766	*syt4/11*	obimac0020553
mynn	obimac0021831	*tnfr*	obimac0032150
nachr	obimac0008351	*tyrbh*	obimac0021237
nkx2	obimac0032473	*vacht*	obimac0013580
nos	obimac0011102	*vat1*	obimac0023625
ob02586	obimac0002586	*vglut*	obimac0012076
ob10569	obimac0010569		

Table S2.2. Reference gene table. Related to Figures 2.2-2.7. Obimac gene identifiers and corresponding gene assignment for all genes described in this study. See Methods for more information about gene assignments.

Appendix A References

S2.1. Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J. Mol. Biol. 430, 2237–2243.

S2.2. Gabler, F., Nam, S.-Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A.N., and Alva, V. (2020). Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. Curr. Protoc.Bioinformatics 72, e108.

S2.3. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

S2.4. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2021). ColabFold - Making protein folding accessible to all. Research Square. 10.21203/rs.3.rs-1032816/v1.

S2.5. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.

S2.6. van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Söding, J., and Steinegger, M. (2022). Foldseek: fast and accurate protein structure search. bioRxiv, 2022.02.07.479398. 10.1101/2022.02.07.479398.

S2.7. Hallgren, J., Tsirigos, K.D., Pedersen, M.D., Armenteros, J.J.A., Marcatili, P., Nielsen, H., Krogh, A., and Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv, 2022.04.08.487609. 10.1101/2022.04.08.487609.

S2.8. Strzalka, W., and Ziemienowicz, A. (2011). Proliferating cell nuclear antigen (PCNA): a key factor in DNA replication and cell cycle regulation. Ann. Bot. 107, 1127–1140.

S2.9. Lawson, A.A., and Barr, R.D. (1987). Acetylcholinesterase in red blood cells. American Journal of Hematology 26, 101–112.

S2.10. Howard, B.V., Macarak, E.J., Gunson, D., and Kefalides, N.A. (1976). Characterization of the collagen synthesized by endothelial cells in culture. Proc. Natl. Acad. Sci. U. S. A. 73, 2361–2364.

S2.11. Martinez-Hernandez, A., Bell, K.P., and Norenberg, M.D. (1977). Glutamine synthetase: glial localization in brain. Science 195, 1356–1358.

S2.12. Cowden, R.R. (1972). Some cytological and cytochemical observations on the leucopoietic organs, the “white bodies,” of Octopus vulgaris. Journal of Invertebrate Pathology 19, 113–119.

S2.13. Freeman, M.R., and Doherty, J. (2006). Glial cell biology in Drosophila and vertebrates. Trends in Neurosciences 29, 82–90.

S2.14. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

APPENDIX B
SUPPORTING INFORMATION FOR CHAPTER III

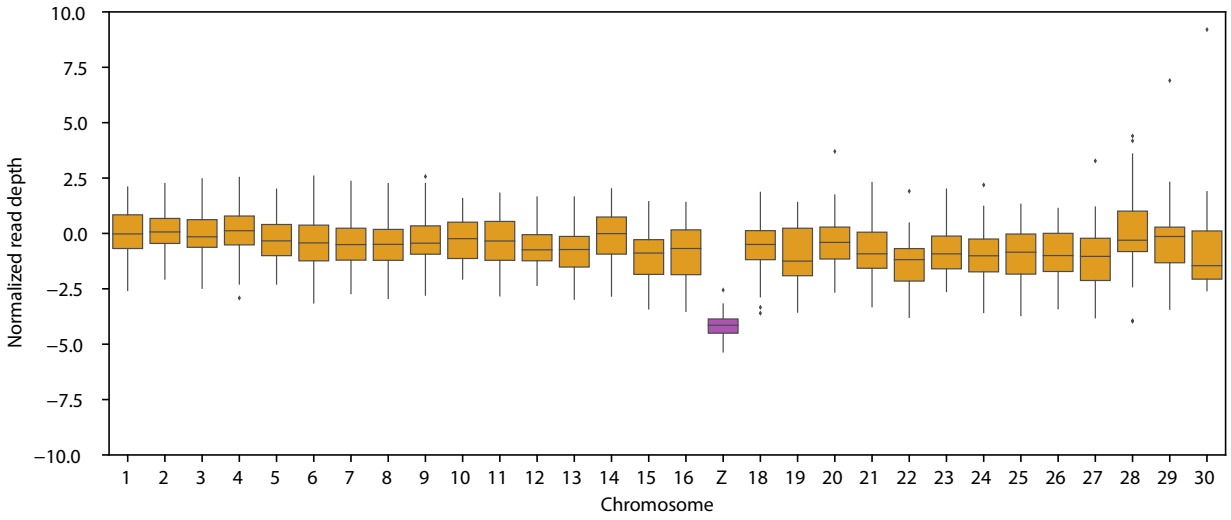

Figure S3.1: Normalized read depth of PacBio HiFi sequencing data across chromosomes from a female O. bimaculoides. Related to Figure 3.1.

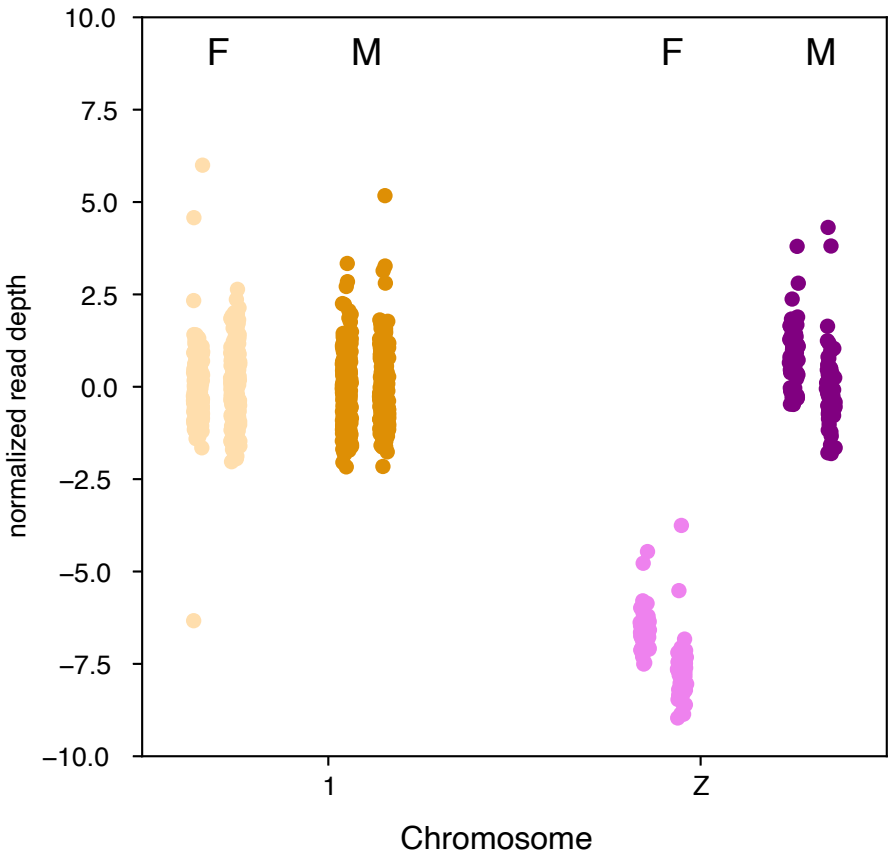

Figure S3.2: Normalized coverage of known sex individuals of O. bimaculoides Related to Figure 3.1 and Table S3.3. Includes two male (darker shades; labelled 'M') and two female (lighter shades; labelled 'F') with short-read sequencing. *O. bimaculoides* individuals at chromosomes 1 and Z are colored to reflect designation in main text. Both female individuals show hemizygous coverage at chromosome Z, whereas both male individuals show the same dosage at chromosome Z as at chromosome 1.

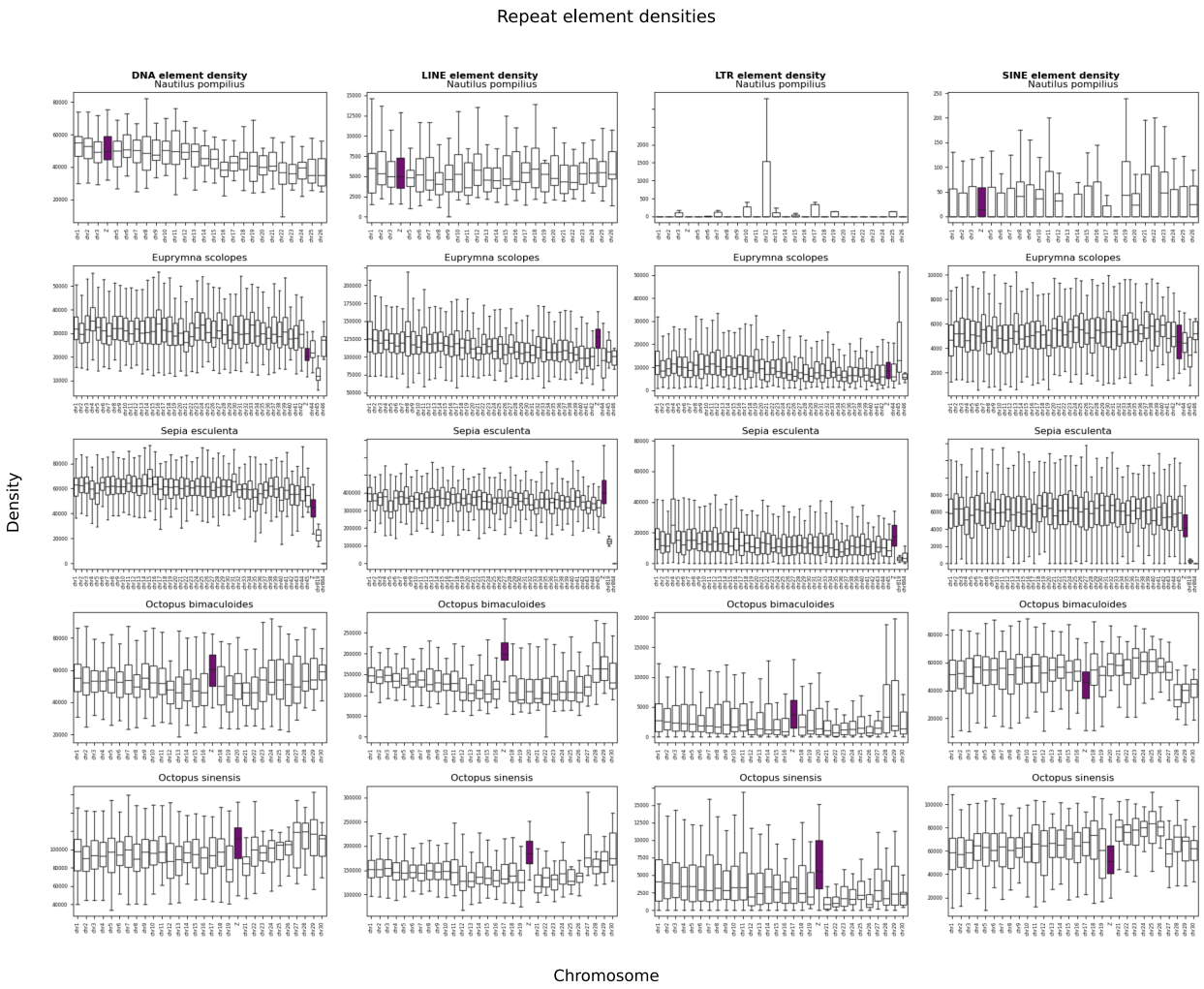

Figure S3.3: Density of repeat elements on chromosomes across cephalopods. Related to Figure 3.2.

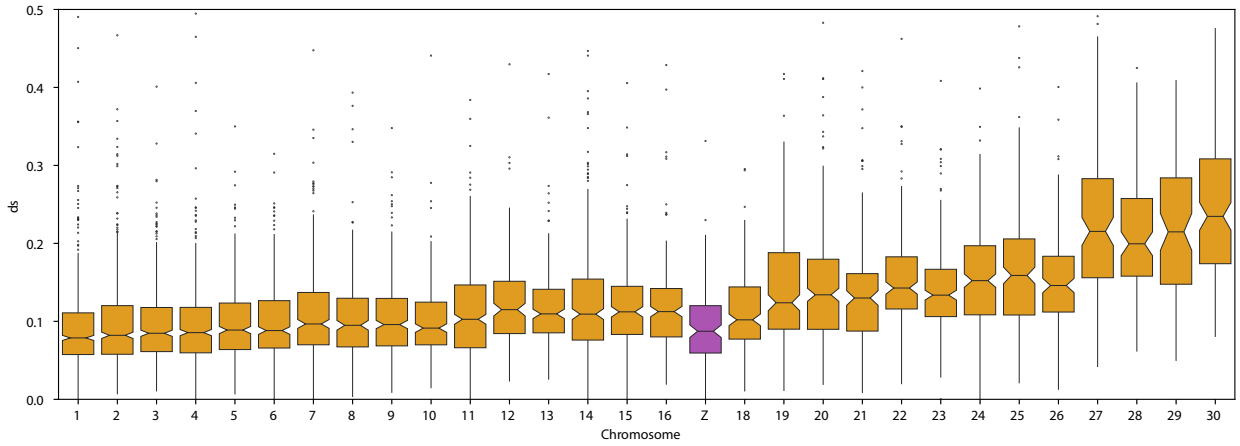

Figure S3.4: Synonymous substitution rate (dS) between *O. bimaculoides* and *O. sinensis* calculated for each gene per chromosome arm. Related to Figure 3.2. The Z chromosome is highlighted in purple whereas autosomes are colored orange.

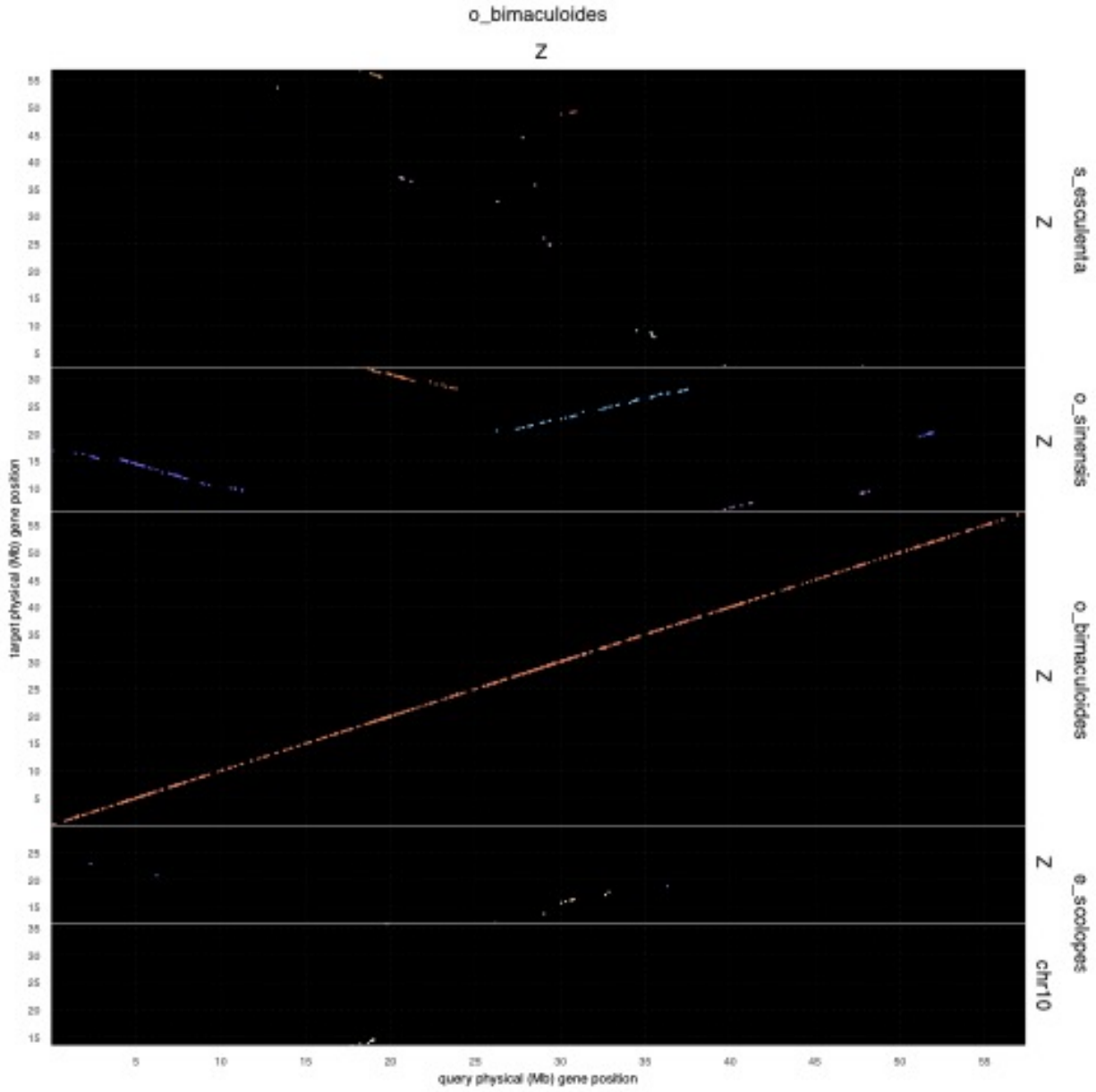

Figure S3.5: Dotplot of syntenic hits between the Z chromosomes of *O. bimaculoides, O. sinensis, E. scolopes,* and *S. esculenta*. Related to Figure 3. The dotplot was generated with GENESPACE v.1.2.3[S3.1] and uses *O. bimaculoides* as the reference species.

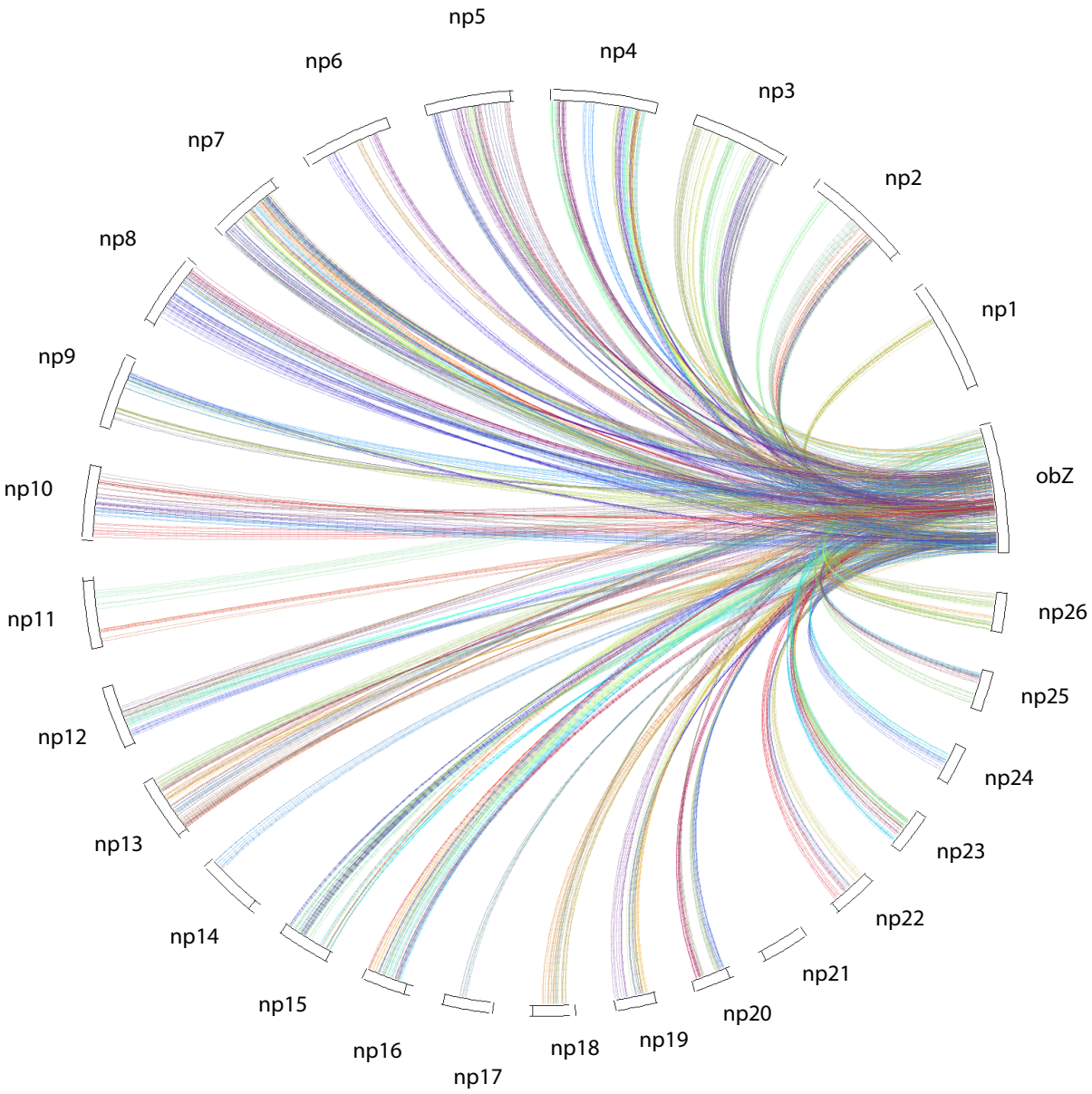

Figure S3.6: Syntenic blocks between the genes of *O. bimaculoides* Z chromosome and all genes of *N. pompilius*. Related to Figure 3.3.

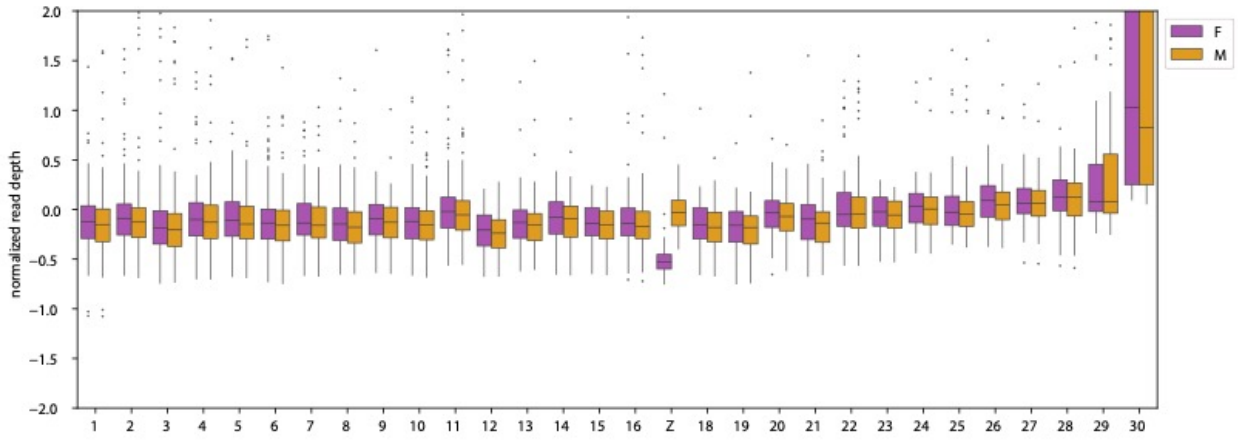

Figure S3.7: Normalized coverage of *O. bimaculatus* male and female short read libraries mapped to our *O. bimaculoides* genome assembly. Related to Figure 3.4. Females have lower coverage than males in the Z chromosome.

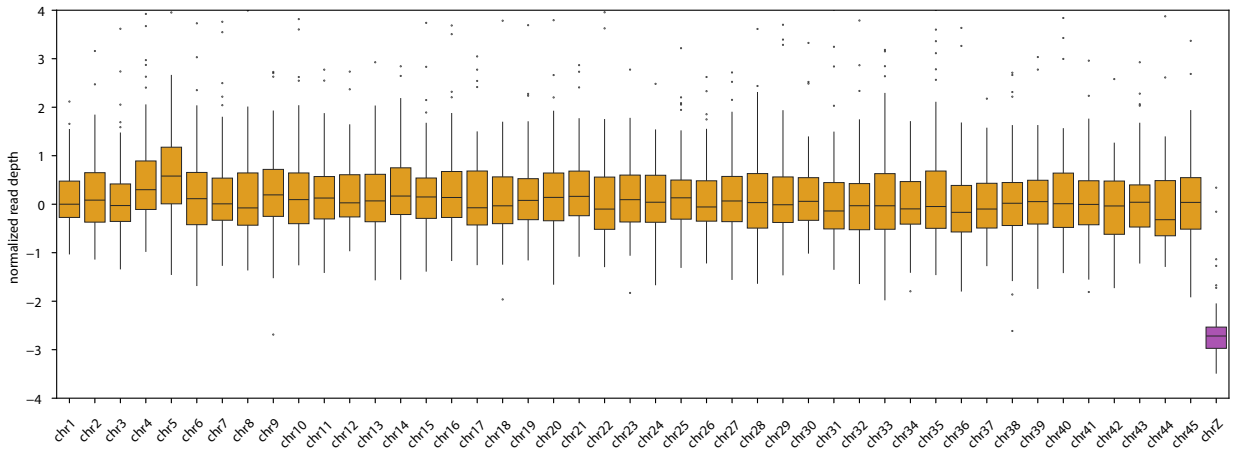

Figure S3.8: Normalized coverage of *S. esculenta* female short read library mapped to the *S. esculenta* genome assembly. Related to Figure 3.4. Data and assembly was generated by the Darwin Tree of Life Project.

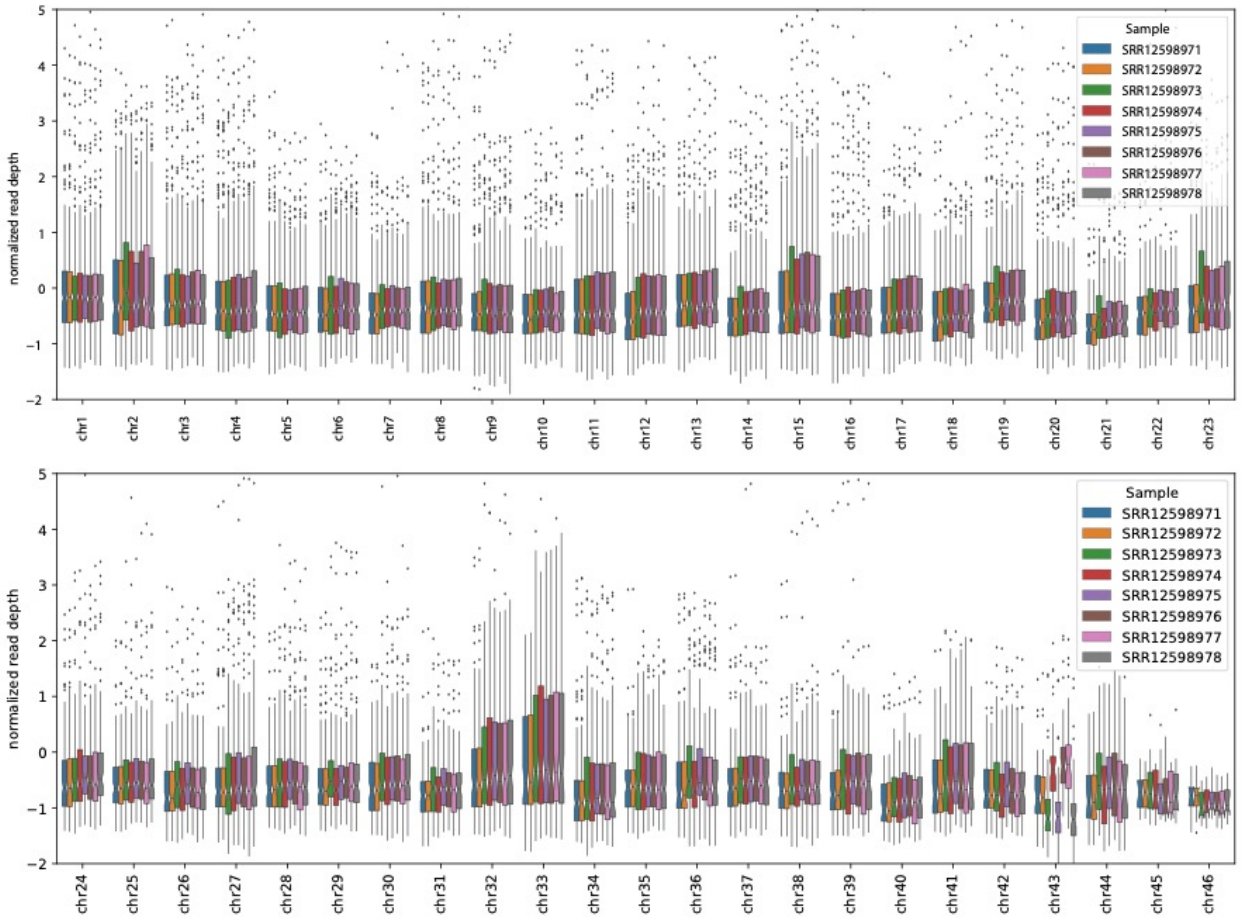

Figure Ss.9: Normalized coverage of unsexed embryo short read libraries mapped to chromosomes in E. scolopes. Related to Figure s.4. Chromosome 43 is the only chromosome with distinct variation in coverage. SRA identifiers for each library are shown in the legend. Data was generated by Schmidbaur et al.[S3.2].

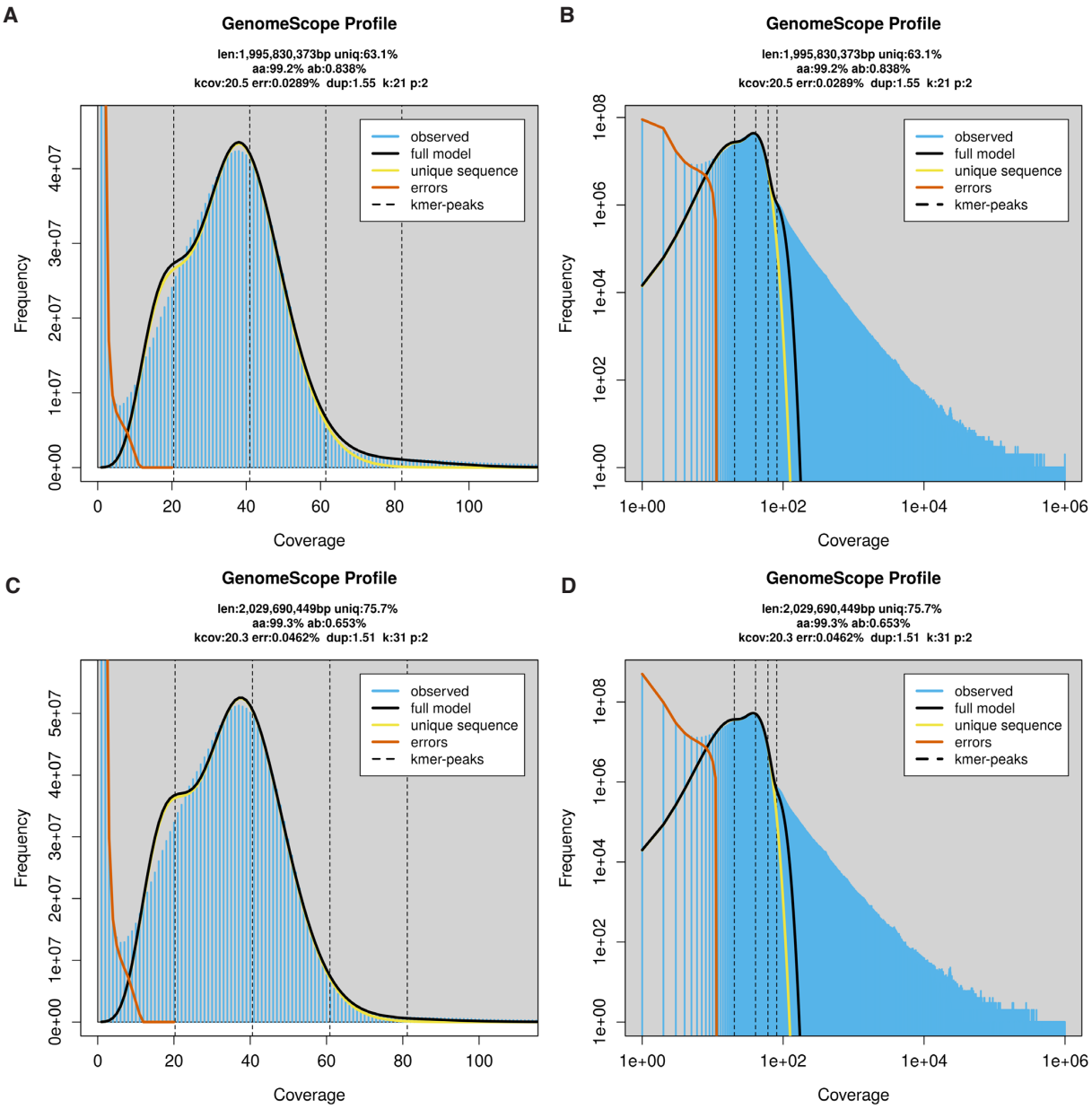

Figure S3.10: K-mer based genome size prediction of O. bimaculoides. Related to Figure 3.1. A and B are 21-mer. C and D are 31-mer. See also Tables S3.1 & S3.2.

	Albertin et al. 2015	Albertin et al. 2022	This Assembly
Size	2.34 Gb	2.37 Gb	2.34 Gb
Number of scaffolds	151,674	145,326	583
Average contig/scaffold length	15,415 bp	6,399 bp	4.02 Mb
Longest scaffold	4.06 Mb	199.87 Mb	202.95 Mb
Contig N50	5,532 bp (n = 97,860)	5,420 bp (n = 100,821)	0.86 Mb (n = 775)
Scaffold N50	0.475 Mb (n = 1333)	96.88 Mb (n = 9)	101.05 Mb (n = 8)
Percent gaps	15.134%	15.016%	0.118%
GC content	36.04%	36.03%	35.79%
Coverage	60x	60x	34x
BUSCO (eukaryote)	C:91.8% [S:90.6%,D:1.2%], F:6.7%, M:1.5%, n : 255	C:92.6% [S:91.4%,D:1.2%], F:6.3%, M:1.1%, n : 255	C:92.2% [S:91.0%,D:1.2%], F:6.7%, M:1.1%, n : 255
BUSCO (metazoan)	C:94.1% [S:93.8%,D:0.3%], F:3.6%, M:2.3%, n : 954	C:94.5% [S:94.1%,D:0.4%], F:3.2%, M:2.3%, n : 954	C:93.9% [S:93.2%,D:0.7%], F:2.6%, M:3.5%, n : 954

Table S3.1: Comparison of Genome Assembly Statistics. Related to Figures 3.1 & S3.10 and Table S3.2. Comparison of the initial genome assembly (Albertin et al. 2015[S3.3]), a subsequent reassembly of those same sequencing data (Albertin et al. 2022[S3.4]) and the present assembly we introduce here.

Number of genes	30655
Number of transcripts	61652
Number of cds	61639
% of genome is cds	1.3%
Number of exons	504491
Mean transcripts per gene	2.0
Mean exons per transcript	8.2
Total gene length (bp)	758146852
Total transcript length (bp)	2799696948
Total cds length (bp)	77744706
Mean gene length (bp)	24731
Mean transcript length (bp)	45411
Mean cds length (bp)	1261
Mean exon length (bp)	279
Mean intron in cds length (bp)	5154
Mean intron in exon length (bp)	6004

Table S3.2: *O. bimaculoides* genome annotation statistics. Related to Figure 3.1 & S3.10 and Table S3.1.

Individual	Sex	Tissue	Data type	Total bases	Read number	Depth	Notes
Ob1	F	Optic lobe	HiFi, PacBio Sequel II 8M	133.4Gb, post-CCS	5.9M	34x	fiveSMRT cells, two libraries
Ob2	Hatchling	Optic lobe, central brain, arm crown, retina	IsoSeq, PacBio Sequel II 8M	10.4Gb	1.6M (64,190 post IsoSeq3)	0.04x	One SMRT cell
Ob3	Hatchling	Arm crown	Hi-C, Illumina NovaSeq 6000	555.8Gb	3.7B	-	One flow cell, two runs
Obmac_F_2	F	Optic lobe	Hi-C, Illumina NovaSeq 6000	90.3Gb	1.24B	71x	-
Obmac_M_3	M	Optic lobe	Hi-C, Illumina NovaSeq 6000	76.2Gb	1.05B	60x	-
Obmac_F_10	F	Gill	Hi-C, Illumina NovaSeq 6000	80.7Gb	1.13B	64x	-
Obmac_M_14	M	Testes	Hi-C, Illumina NovaSeq 6000	66.0Gb	924M	52x	-

Table S3.3: Statistics of raw sequencing data collected from *O. bimaculoides* samples. Related to Figures 3.1, S3.1, & S3.2, and Tables S3.1 & S3.2.

Individual	Sex	Tissue	Data type	Total bases	Number of reads	Depth
Oblat_F_2	F	Optic lobe	Illumina NovaSeq 6000	94.0Gb	669M	35x
Oblat_M_3	M	Optic lobe	Illumina NovaSeq 6000	92.6Gb	664M	34x
Oblat_F_10	F	Optic lobe	Illumina NovaSeq 6000	111.4Gb	764M	42x
Oblat_M_4	M	Optic lobe	Illumina NovaSeq 6000	111.6GB	779M	42x

Table S3.4: Statistics of raw sequencing data collected form *O. bimaculatus* samples. Related to Figure 3.4.

Species	Repeat Type	p-value
Sepia esculenta	DNA	**4.06E-20**
Euprymna scolopes	DNA	**2.91E-17**
Octopus bimaculoides	DNA	1.34E-04
Octopus sinensis	DNA	3.04E-03
Octopus bimaculoides	LINE	**9.90E-25**
Octopus sinensis	LINE	**5.71E-09**
Sepia esculenta	LINE	**6.21E-05**
Euprymna scolopes	LINE	**2.34E-04**
Octopus bimaculoides	LTR	**1.23E-04**
Octopus sinensis	LTR	1.06E-03
Sepia esculenta	LTR	**3.64E-03**
Euprymna scolopes	LTR	3.05E-01
Octopus bimaculoides	SINE	**3.66E-06**
Sepia esculenta	SINE	**7.28E-06**
Octopus sinensis	SINE	1.45E-03
Euprymna scolopes	SINE	1.90E-02

Table S3.5: Mann–Whitney U tests by repeat element type. Related to Figure 3.2. Tested for a significant difference between the sex chromosome element density against all autosomes. Bold indicates significance after Bonferroni correction.

Octo Gene ID	Human Gene ID	Human Gene Symbol	E-val	Function	Human Reproductive mRNA Expression?	Human Reproductive Protein Expression?	Found on Nautilus Z?
obimac 0008617.1	NP 001121055.1	ING4	1.00E-109	Tumor suppressor protein that contains a PHD-finger	ovary, uterus, placenta, prostate, testis	ovary (fetal), testis, testis (fetal)	0
obimac 0008634.1	NP 001182257.1	RAB9A	2.00E-92	GDP binding activity; GTP binding activity; and GTPase activity	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	1
obimac 0008637.2	NP 001007469.1	SMARCB1	0.00E+00	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin	ovary, uterus, placenta, prostate, testis	uterus, ovary, ovary (fetal), testis, testis (fetal)	0
obimac 0008638.1	no hit	-	-	-	-	-	0
obimac 0008640.1	NP 006310.1	CDIPT	8.00E-81	Catalyzes the biosynthesis of phosphatidylinositol	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	0
obimac 0008642.1	no hit	-	-	-	-	-	1
obimac 0008643.1	NP 005207.3	DDOST	8.00E-128	Subunit of the oligosaccharyl transferase (OST) complex	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal), seminal vesicle	0
obimac 0008647.1	NP 003123.2	SRM	2.00E-114	Spermidine Synthase	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal), seminal vesicle	0
obimac 0008648.1	NP 001011547.2	SLC5A9	0.00E+00	Electrogenic Na(+)-coupled sugar symporter	ovary, uterus, prostate, testis	none	1
obimac 0008657.1	NP 004152.1	RAB1A	9.00E-134	Ras superfamily of GTPases	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal), seminal vesicle	1
obimac 0008668.1	NP 001287785.1	FOSL1	2.00E-13	FOS Like 1, AP-1 Transcription Factor Subunit	ovary, uterus, prostate, testis	ovary	0
obimac 0008926.1	AAH14937.1	FHIP1A	3.00E-129	Protein localization to perinuclear region of cytoplasm	ovary, uterus, prostate, testis	testis (fetal)	1
obimac 0008950.6	XP 005257828.1	SPAG9	0.00E+00	Sperm Associated Antigen; cancer testis antigen gene family	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	1
obimac 0008956.1	no hit	-	-	-	-	-	1
obimac 0008962.2	NP 002755.1	PRPS1	0.00E+00	Catalyzes the phosphoribosylation of ribose 5-phosphate to 5-phosphoribosyl-1-pyrophosphate	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	0
obimac 0008963.1	NP 005330.1	UBE2K	9.00E-89	Ubiquitin-conjugating enzyme	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal), seminal vesicle	0
obimac 0008964.1	NP 001073308.1	OCIAD1	9.00E-22	OCIA Domain Containing 1; Maintains stem cell potency	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	0
obimac 0008989.8	NP 001381782.1	KLC1	0.00E+00	Kinesin Light Chain 1	ovary, uterus, placenta, prostate, testis	-	1
obimac 0008993.1	NP 001960.2	EIF5	8.00E-147	Eukaryotic translation initiation factor-5	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	1

Table S3.6: Table of blast results from orthologous region genes of chromosome Z to humans. Related to Figure 3.3.

Appendix B References

S.3.1. Lovell, J.T., Sreedasyam, A., Schranz, M.E., Wilson, M., Carlson, J.W., Harkess, A., Emms, D., Goodstein, D.M., and Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. eLife 11, e78526. https://doi.org/10.7554/eLife.78526.

S3.2. Schmidbaur, H., Kawaguchi, A., Clarence, T., Fu, X., Hoang, O.P., Zimmermann, B., Ritschard, E.A., Weissenbacher, A., Foster, J.S., Nyholm, S.V., et al. (2022). Emergence of novel cephalopod gene regulation and expression through large-scale genome reorganization. Nat Commun 13, 2172. https://doi.org/10.1038/s41467-022- 29694-7.

S3.3. Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., and Rokhsar, D.S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524, 220–224. https://doi.org/10.1038/nature14668.

S3.4. Albertin, C.B., Medina-Ruiz, S., Mitros, T., Schmidbaur, H., Sanchez, G., Wang, Z.Y., Grimwood, J., Rosenthal, J.J.C., Ragsdale, C.W., Simakov, O., et al. (2022). Genome and transcriptome mechanisms driving cephalopod evolution. Nat Commun 13, 2427. https://doi.org/10.1038/s41467-022-29748-w.

APPENDIX C
SUPPORTING INFORMATION FOR CHAPTER IV

Species	Individual	Sex	Tissue	Percent mapped	Number of reads	Depth	Previously published?
O. bimaculatus	Oblat_F_2	F	Optic lobe	99.14%	669M	35x	Yes
O. bimaculatus	Oblat_F_5	F	Optic lobe	99.39%	738M	35x	No
O. bimaculatus	Oblat_F_6	F	Optic lobe	99.04%	719M	40x	No
O. bimaculatus	Oblat_F_7	M	Optic lobe	99.00%	703M	39x	No
O. bimaculatus	Oblat_F_8	M	Optic lobe	99.42%	425M	21x	No
O. bimaculatus	Oblat_F_9	F	Optic lobe	99.23%	920M	49x	No
O. bimaculatus	Oblat_F_10	F	Optic lobe	99.05%	764M	42x	Yes
O. bimaculatus	Oblat_F_11	F	Optic lobe	99.07%	706M	36x	No
O. bimaculatus	Oblat_M_01	M	Optic lobe	98.78%	664M	34x	No
O. bimaculatus	Oblat_M_3	M	Optic lobe	99.24%	664M	34x	Yes
O. bimaculatus	Oblat_M_4	M	Optic lobe	99.17%	779M	42x	Yes
O. bimaculatus	Oblat_M_12	M	Optic lobe	99.02%	664M	33x	No
O. bimaculatus	Oblat_M_13	M	Optic lobe	99.15%	664M	48x	No
O. bimaculoides	Obmac_F_2	F	Optic lobe	99.19%	1.24B	53x	No
O. bimaculoides	Obmac_F_4	F	Arm, gill	99.25%	976M	37x	No
O. bimaculoides	Obmac_F_10	F	Gill	99.21%	1.13B	44x	No
O. bimaculoides	Obmac_M_3	M	Optic lobe	99.20%	1.05B	43x	No
O. bimaculoides	Obmac_M_8	M	Arm, gill, optic lobe	99.11%	913M	39x	No
O. bimaculoides	Obmac_M_14	M	testes	99.25%	923M	37x	No

Table S1: Raw sequencing statistics for re-sequencing dataset of octopus. Total reads and percent mapped were calculated by the SNPArcher pipeline. Depth per individual was calculated by mosdepth. Datasets that were published previously were used as coverage data in Coffing et al. 2025.

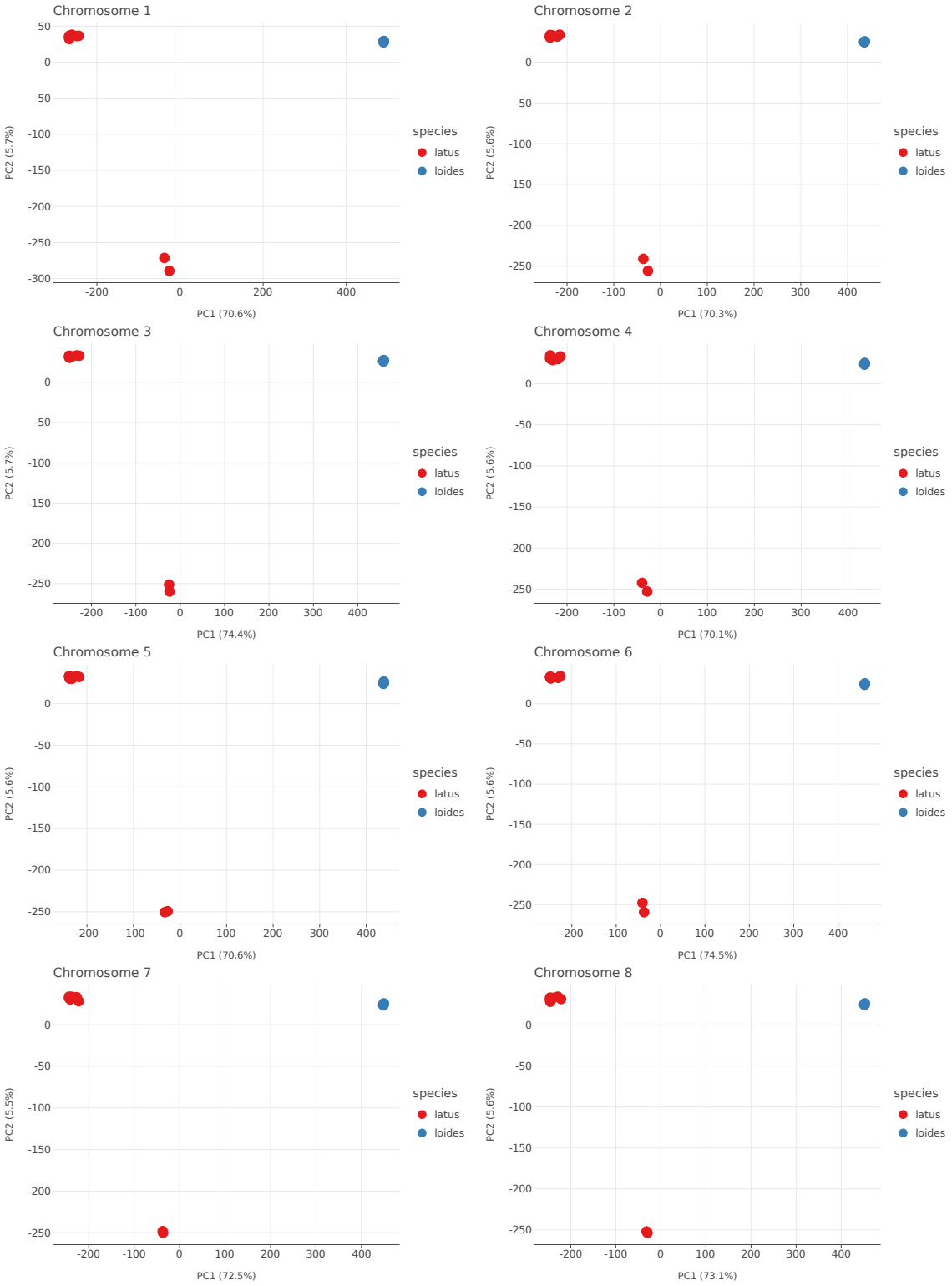

Figure S4.1: PCA of chromosomes 1 – 8 of *O. bimaculoides* and *O. bimaculatus* individuals.

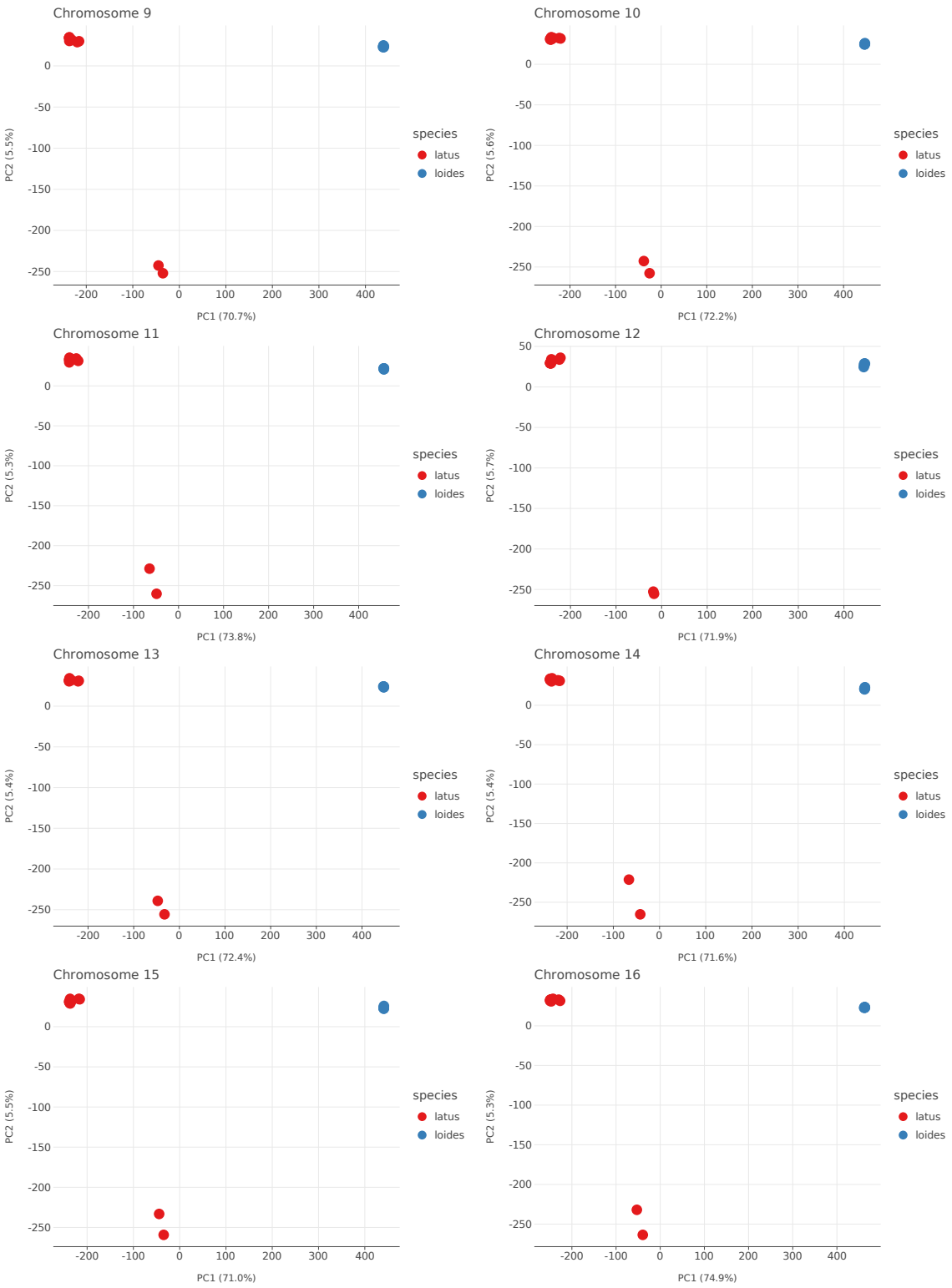

Figure S4.2: PCA of chromosomes 9 – 16 of *O. bimaculoides* and *O. bimaculatus* individuals.

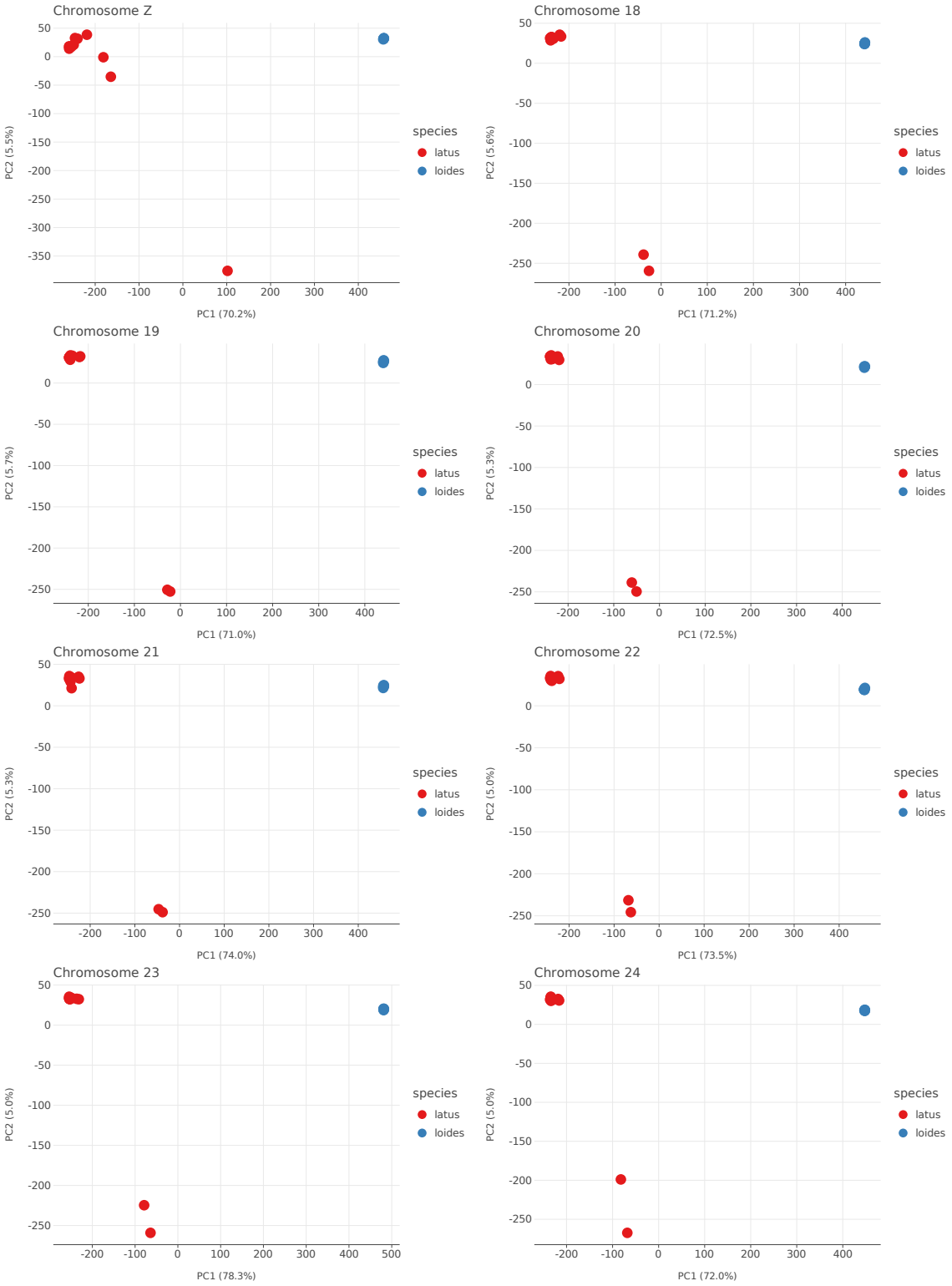

Figure S4.3: PCA of chromosomes Z and 18 – 24 of *O. bimaculoides* and *O. bimaculatus* individuals.

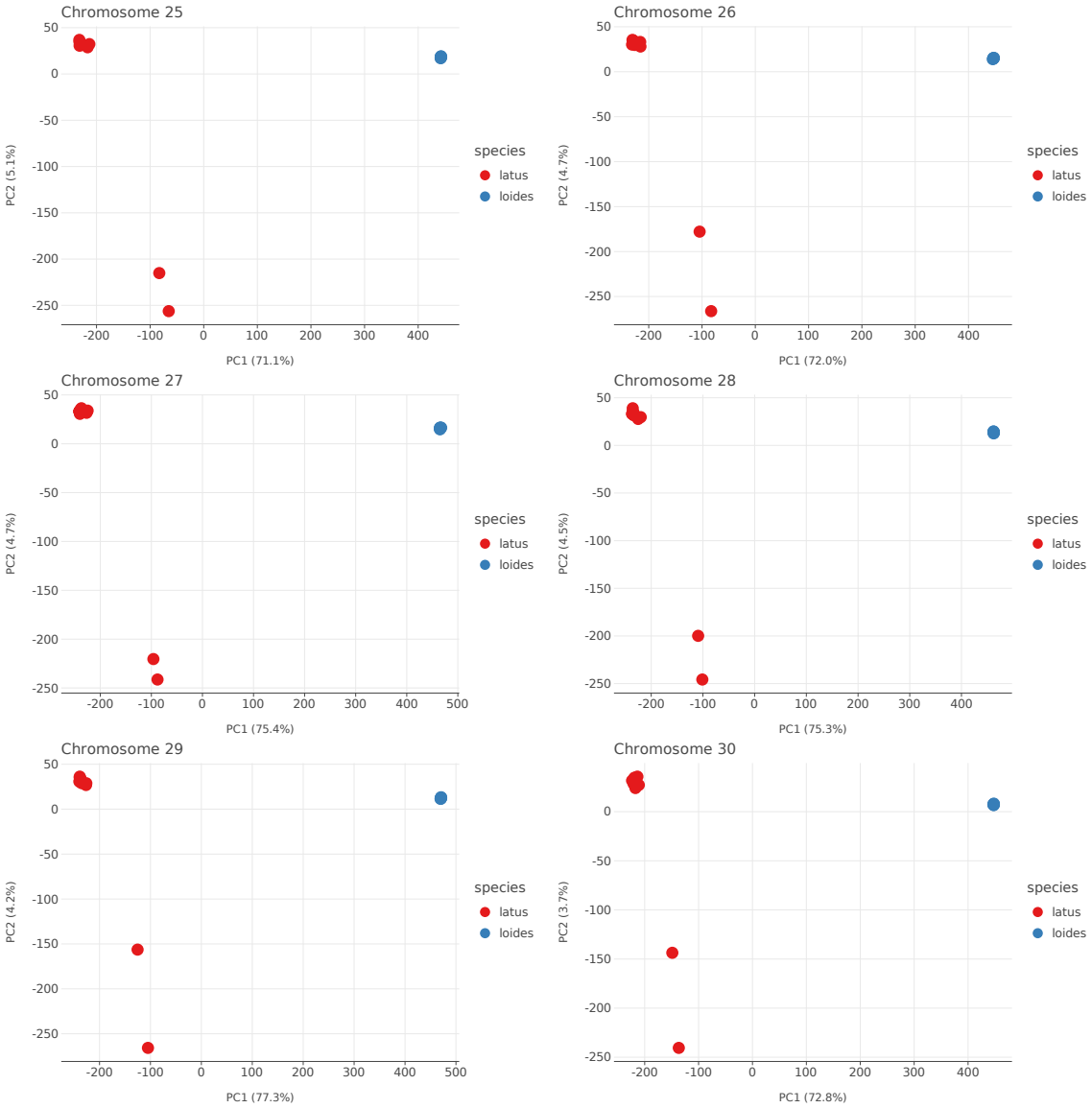

Figure S4.3: PCA of chromosomes 25 – 30 of *O. bimaculoides* and *O. bimaculatus* individuals.

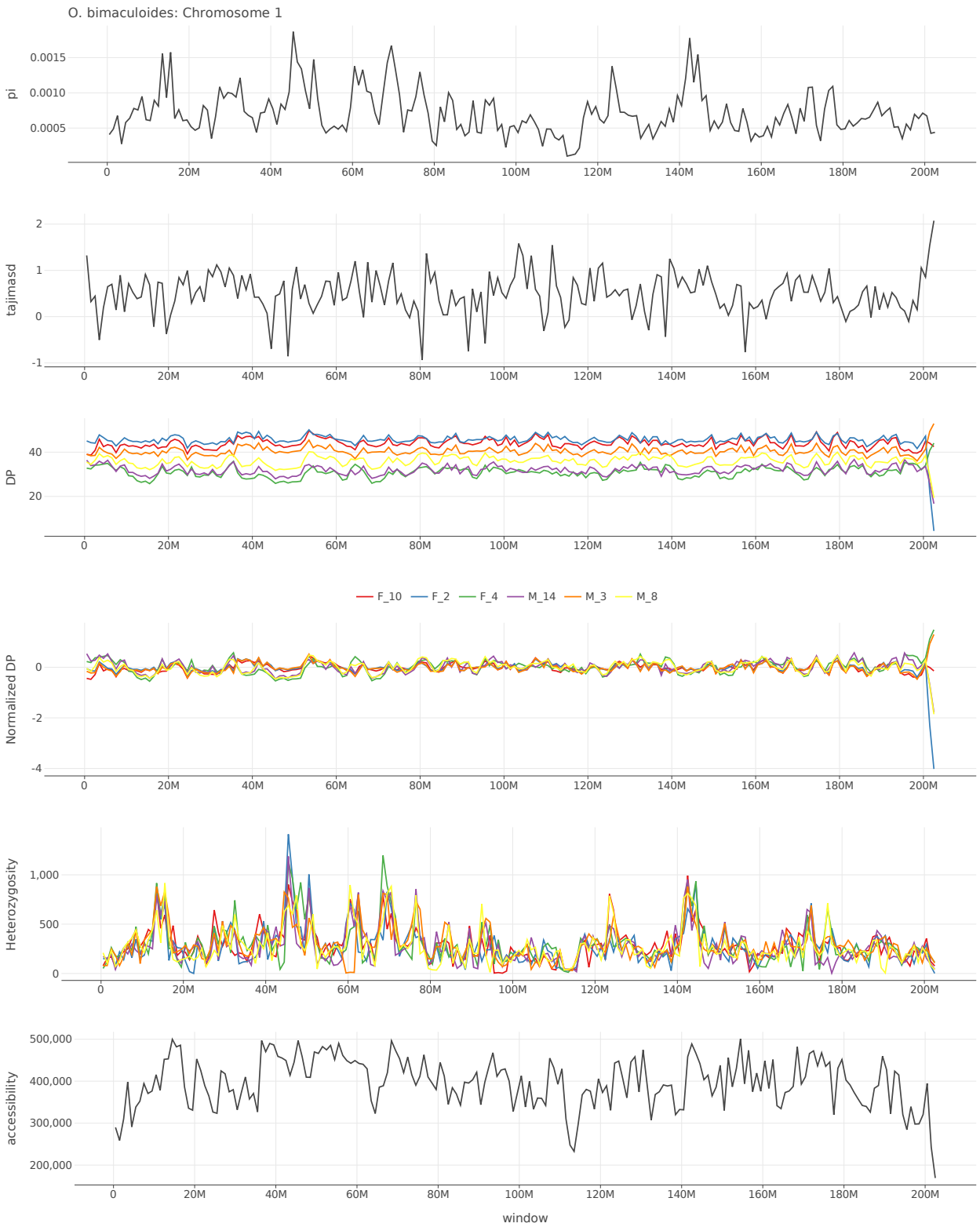

Figure S4.4: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 1 of *O. bimaculoides*.

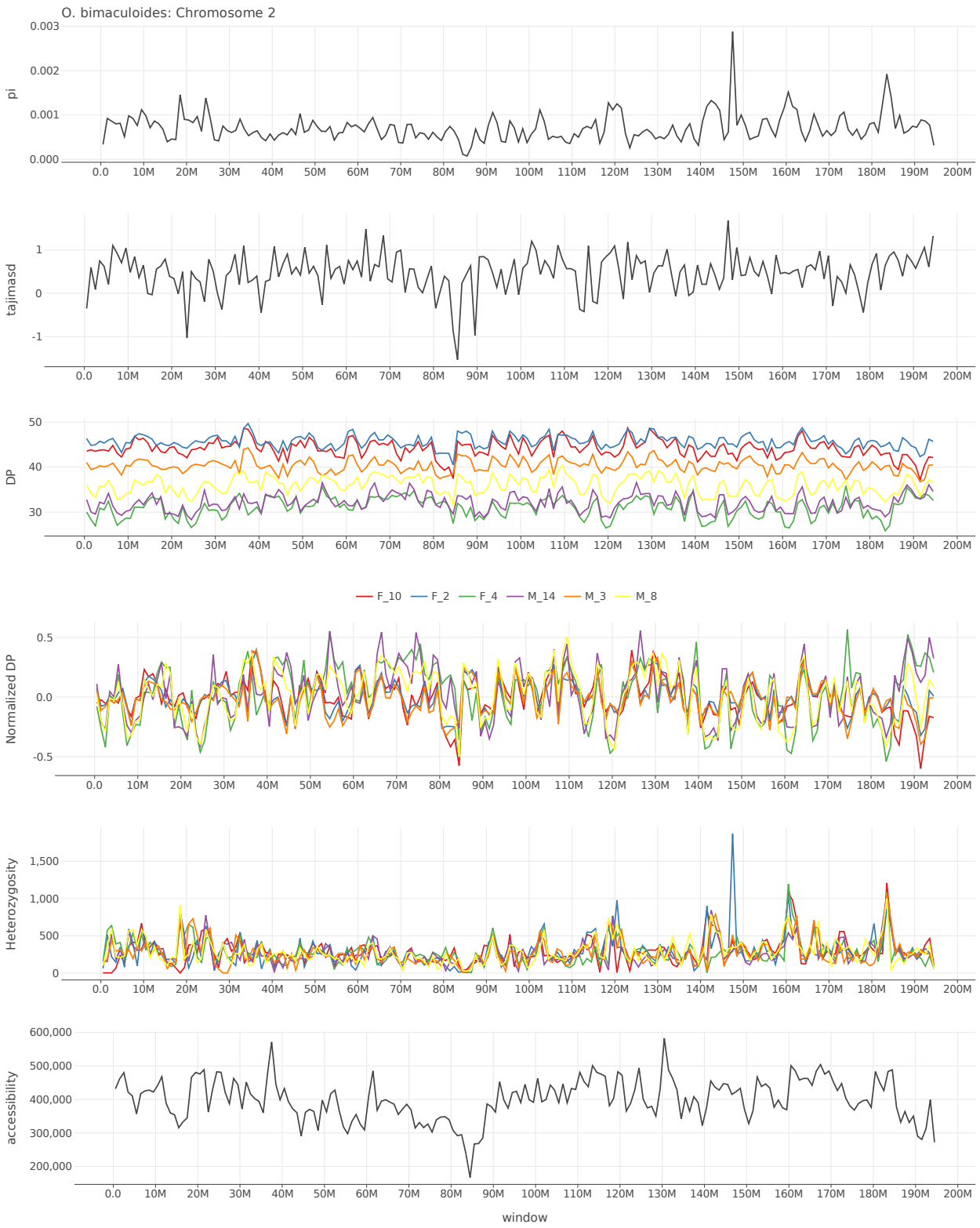

Figure S4.5: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 2 of *O. bimaculoides*.

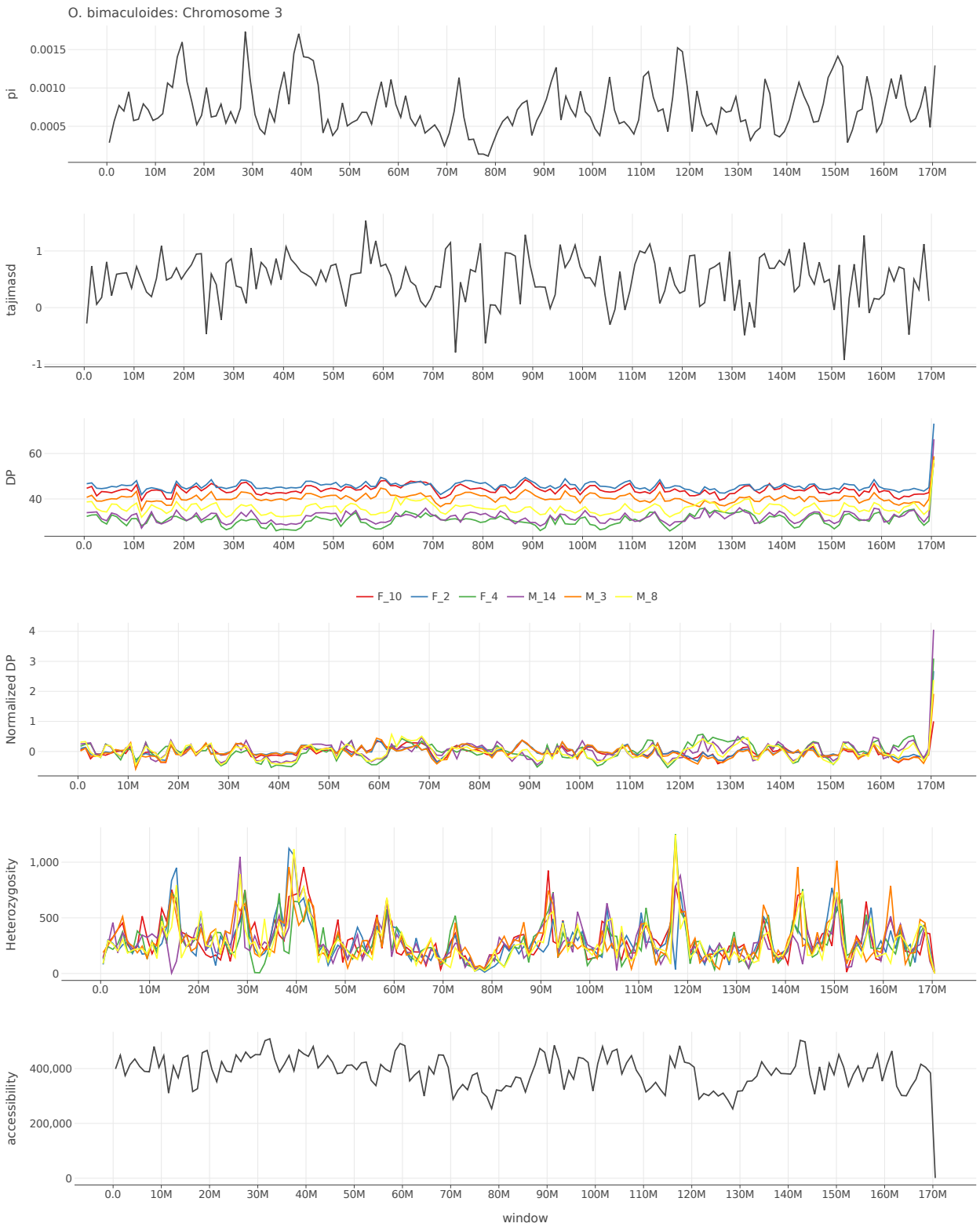

Figure S4.5: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 3 of *O. bimaculoides.*

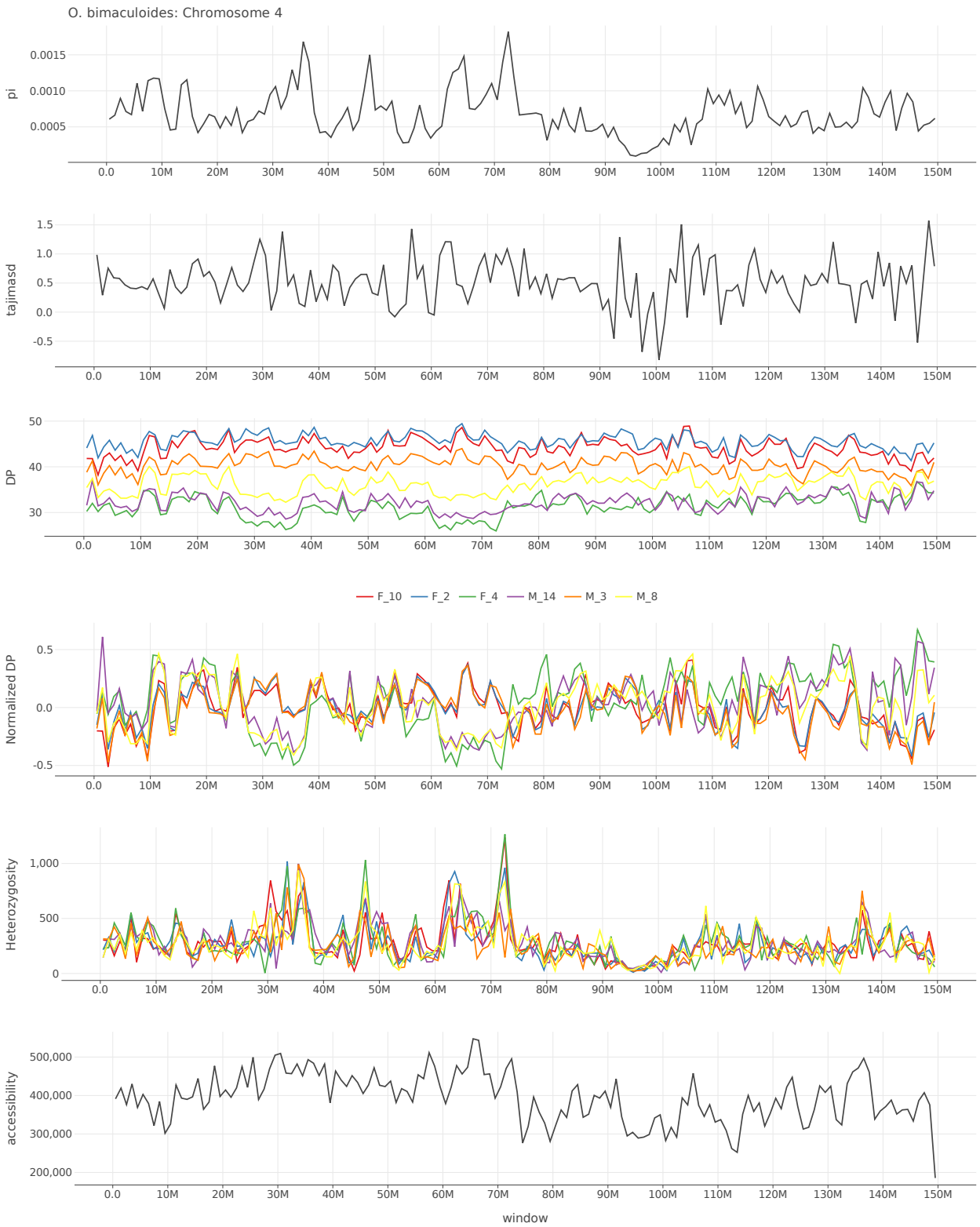

Figure S4.6: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 4 of *O. bimaculoides.*

Figure S4.7: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 5 of *O. bimaculoides*.

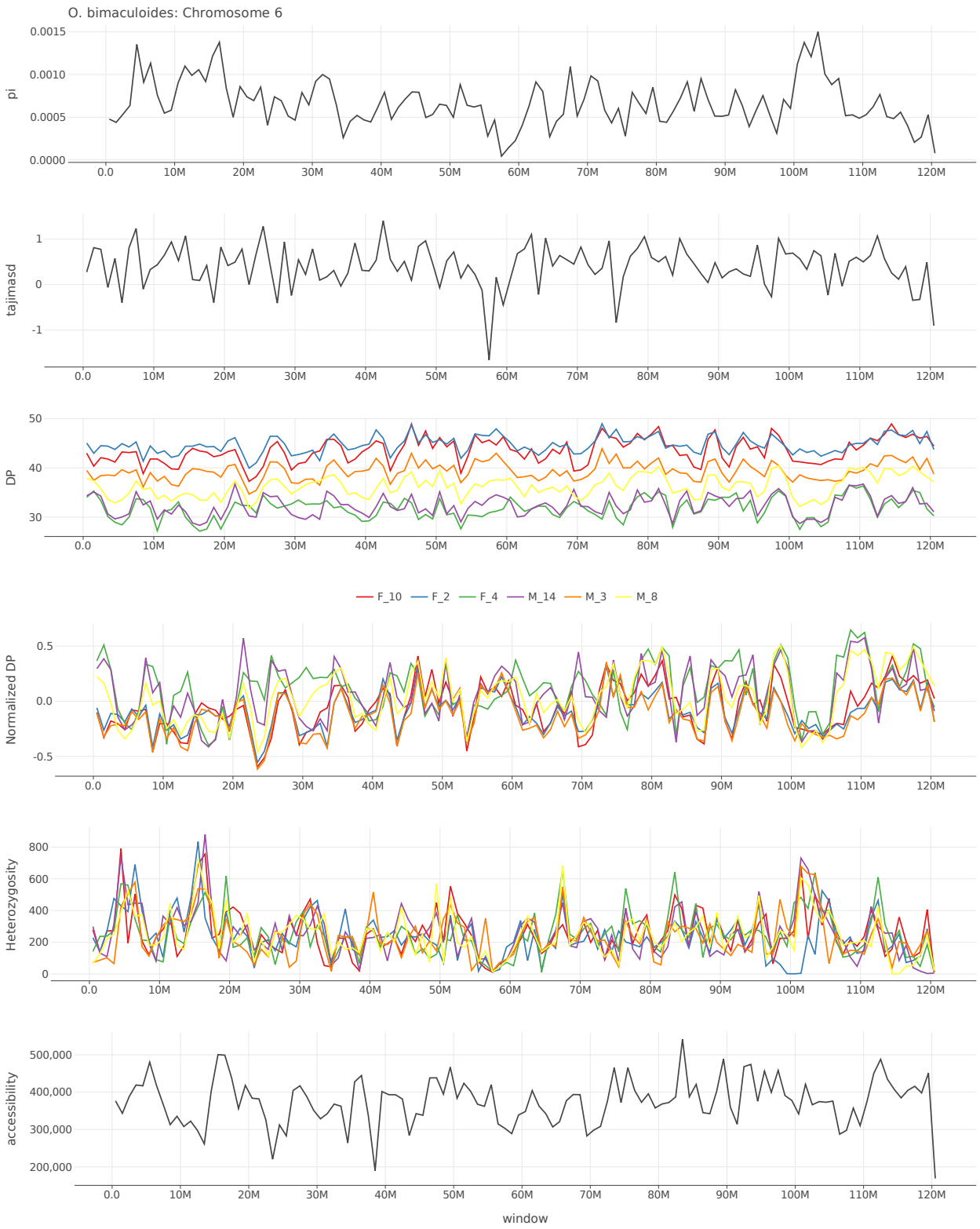

Figure S4.8: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 6 of *O. bimaculoides*.

Figure S4.9: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 7 of *O. bimaculoides*.

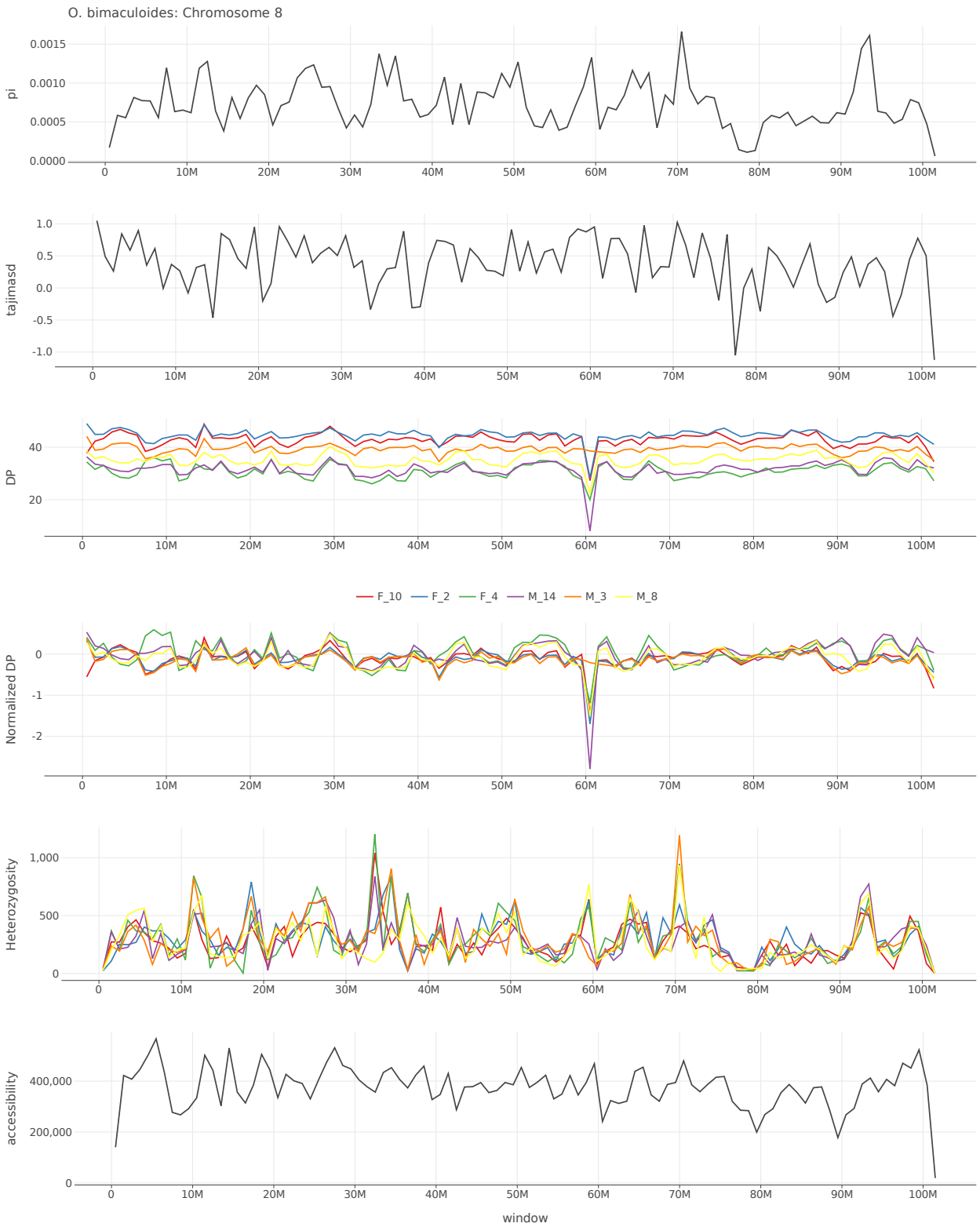

Figure S4.10: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 8 of *O. bimaculoides.*

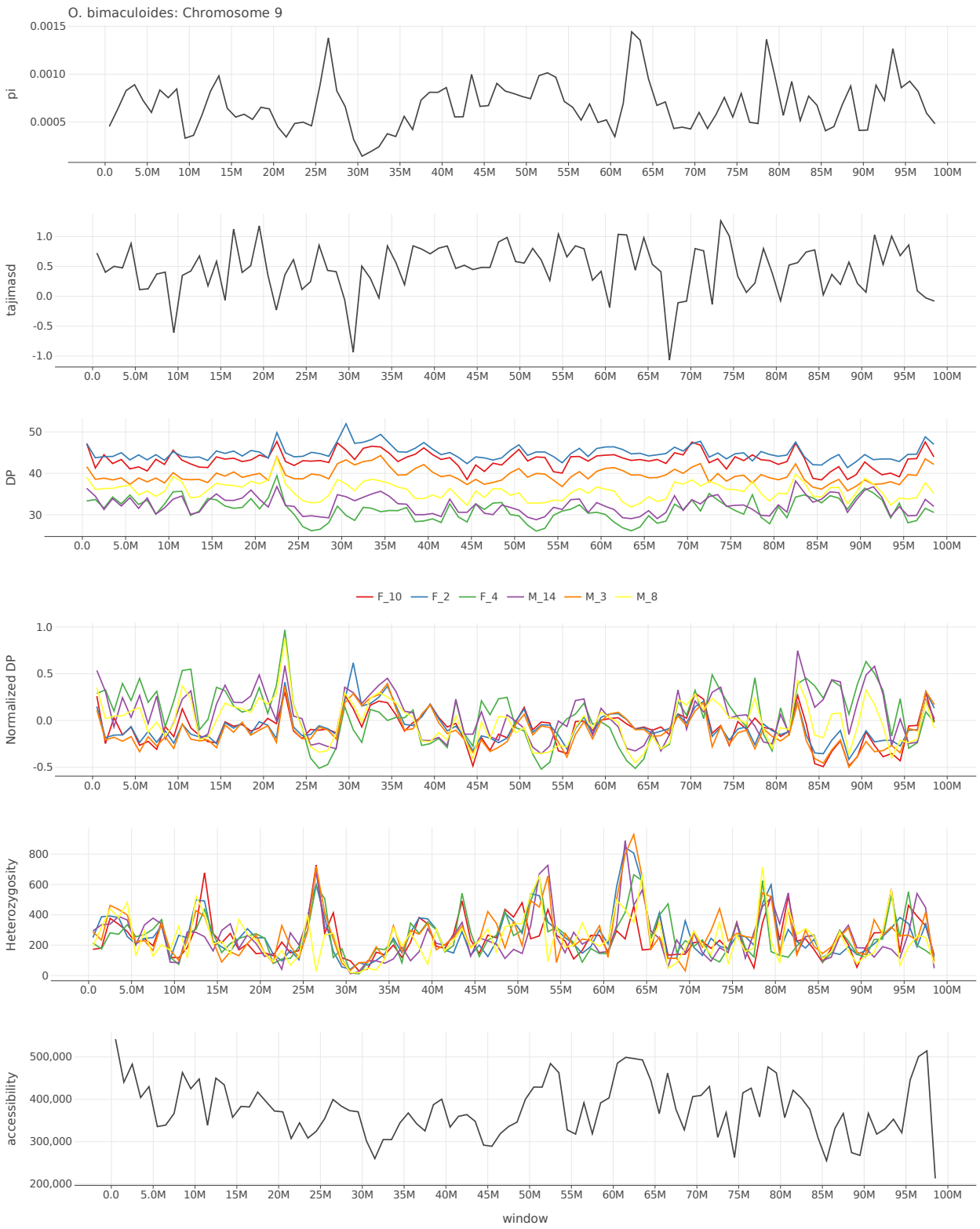

Figure S4.11: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 9 of *O. bimaculoides*.

Figure S4.12: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 10 of *O. bimaculoides*.

Figure S4.13: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 11 of *O. bimaculoides*.

Figure S4.14: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 12 of *O. bimaculoides*.

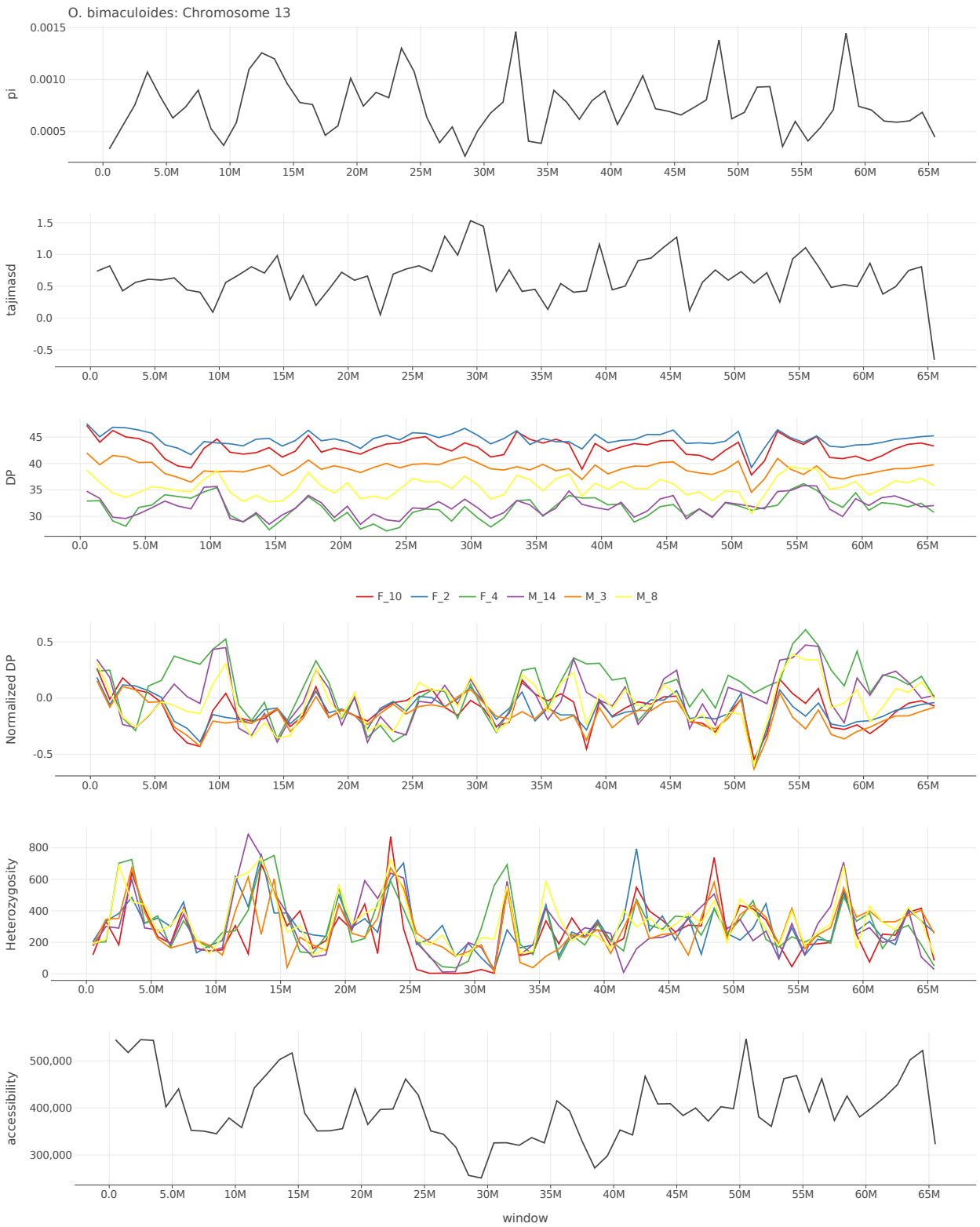

Figure S4.15: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 13 of *O. bimaculoides*.

Figure S4.16: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 14 of *O. bimaculoides.*

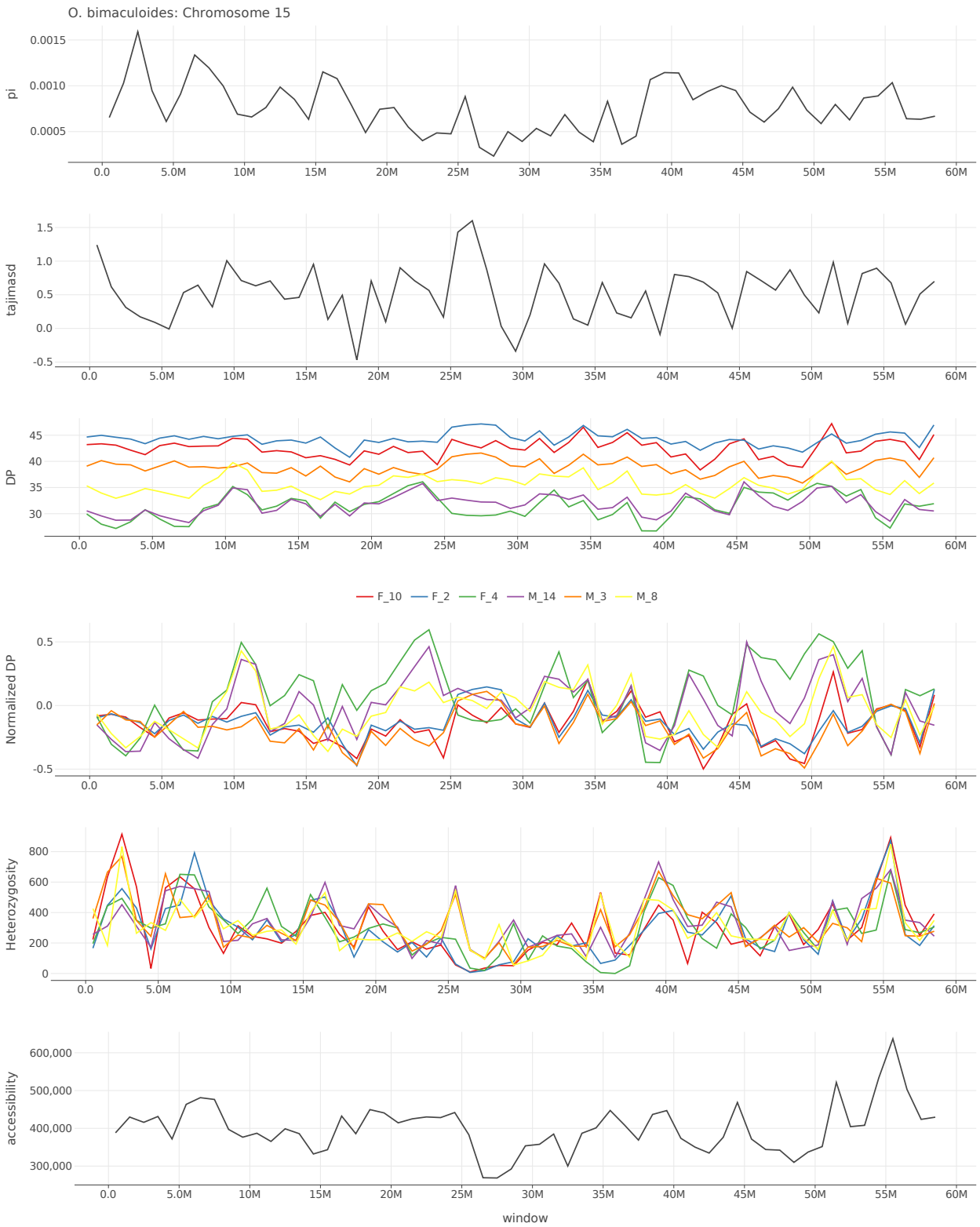

Figure S4.17: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 15 of *O. bimaculoides*.

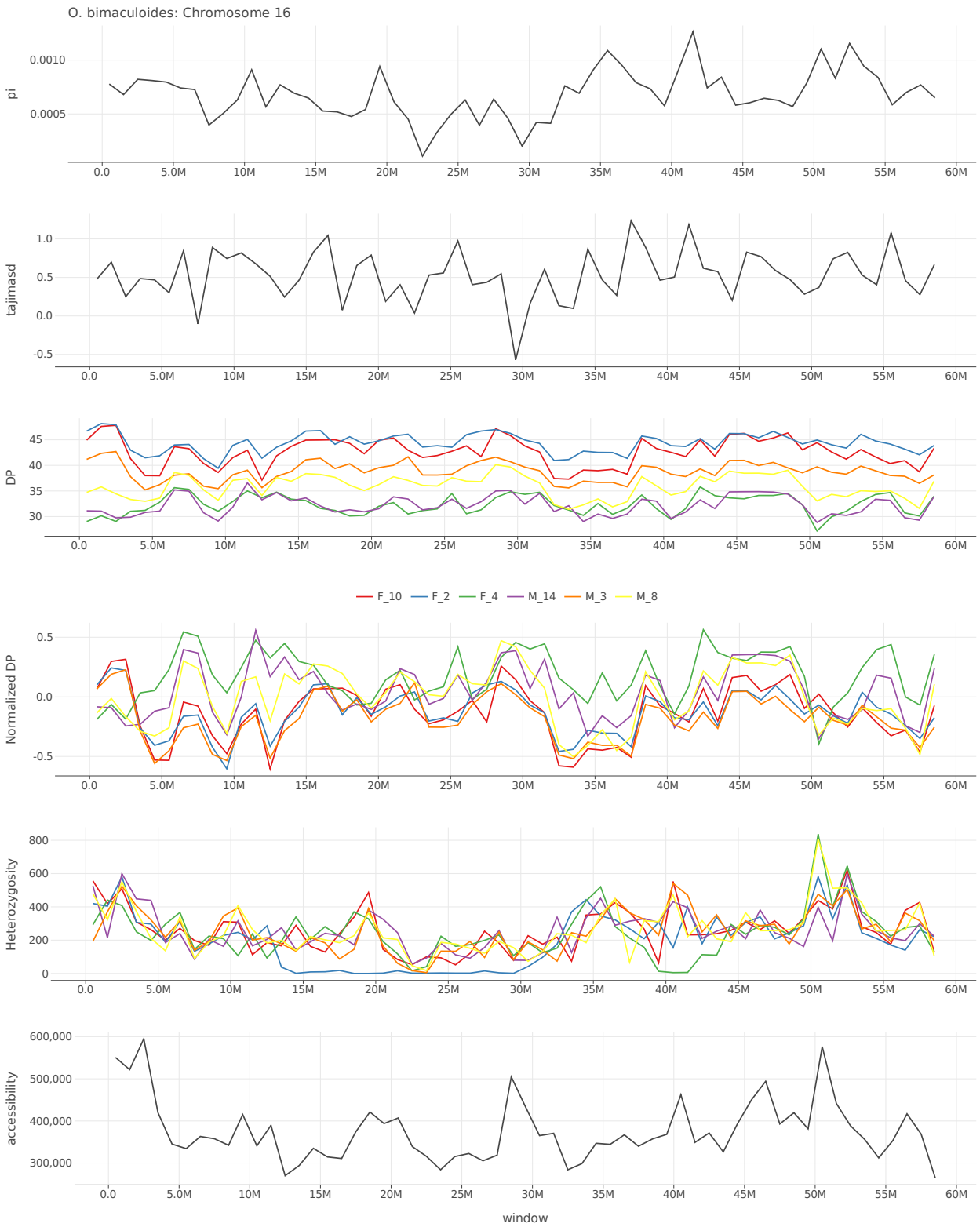

Figure S4.18: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 16 of *O. bimaculoides*.

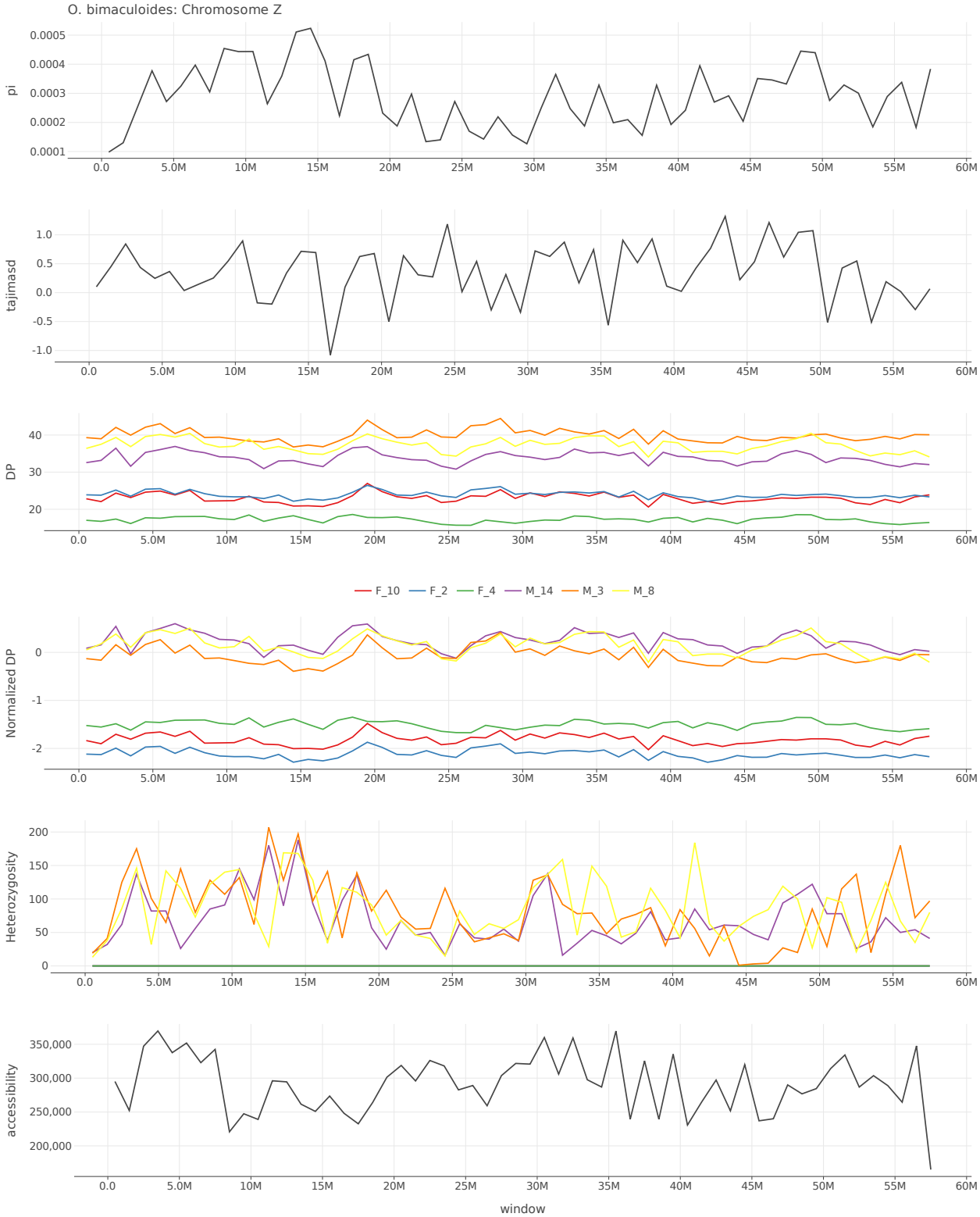

Figure S4.19: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome Z of *O. bimaculoides.*

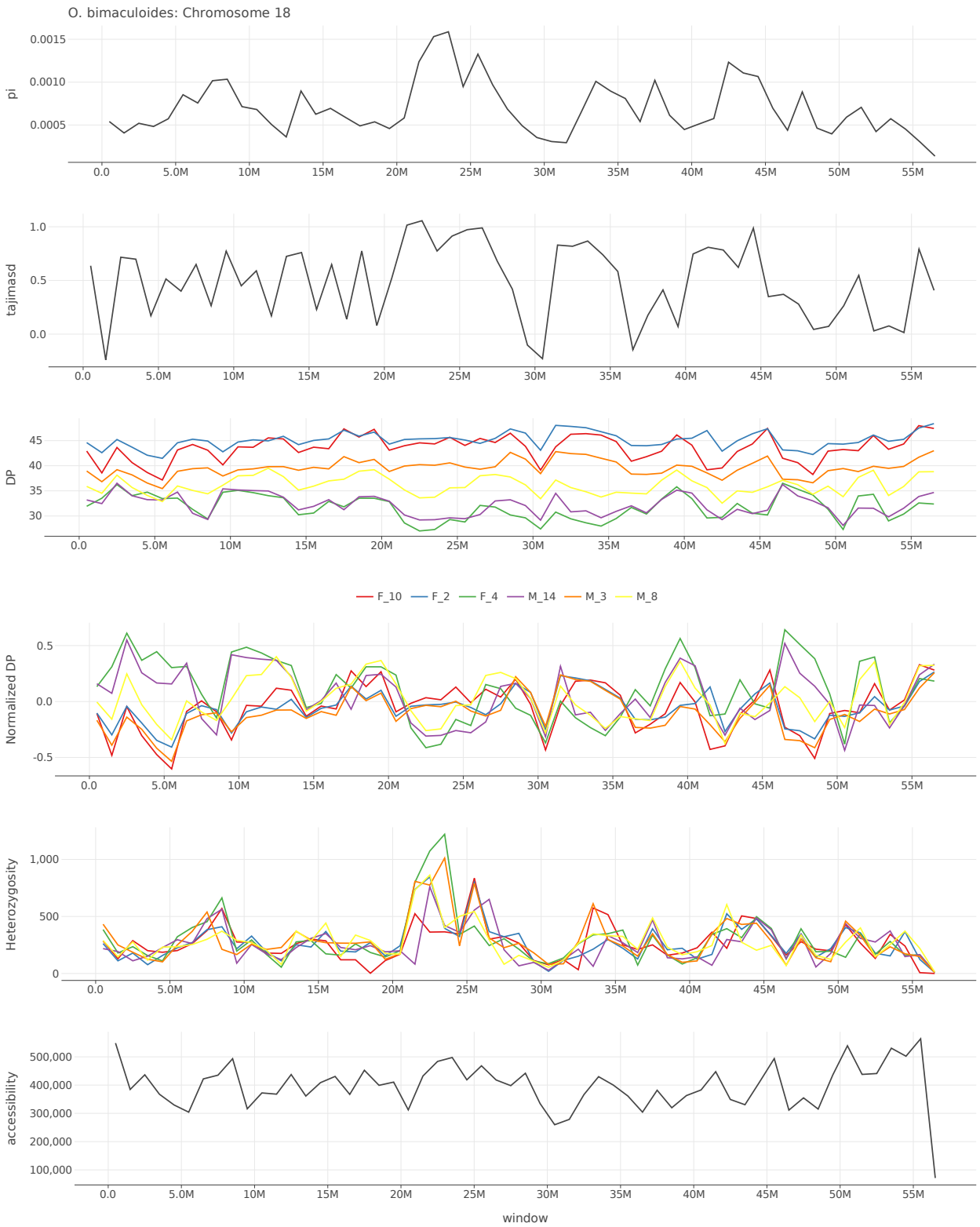

Figure S4.20: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 18 of *O. bimaculoides*.

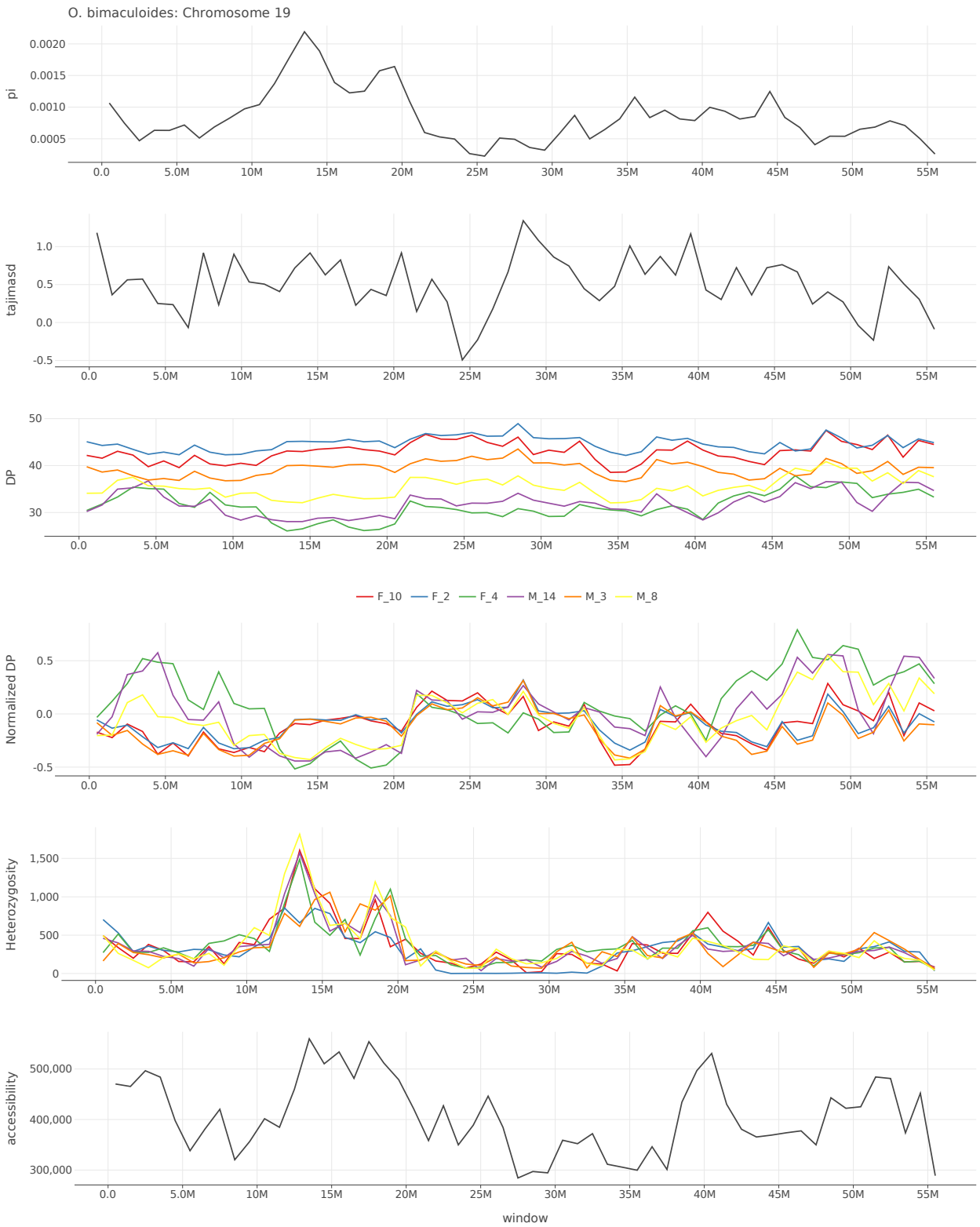

Figure S4.21: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 19 of *O. bimaculoides*.

Figure S4.22: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 20 of *O. bimaculoides*.

Figure S4.23: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 21 of *O. bimaculoides*.

Figure S4.24: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 22 of *O. bimaculoides*.

Figure S4.25: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 23 of *O. bimaculoides*.

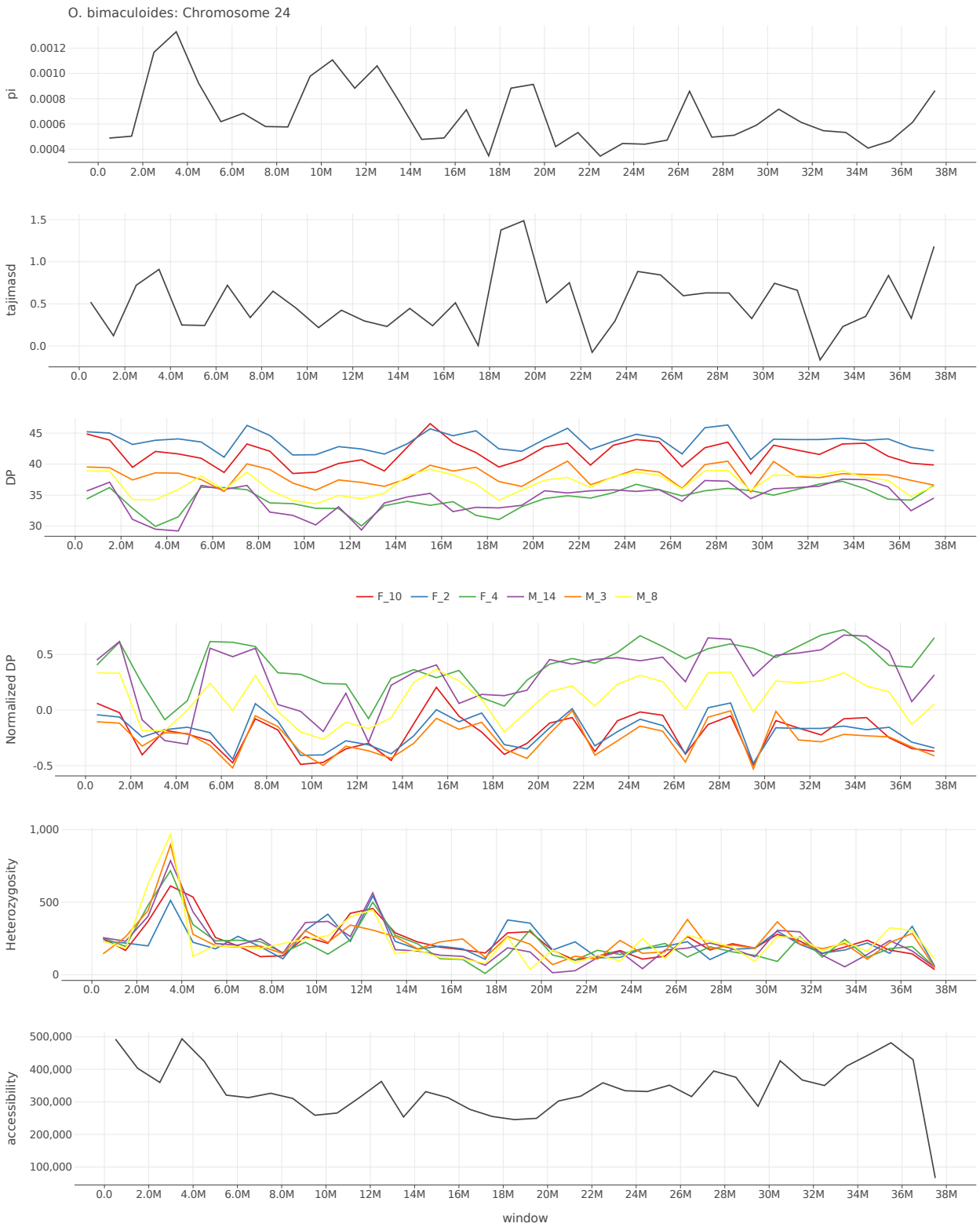

Figure S4.26: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 24 of *O. bimaculoides*.

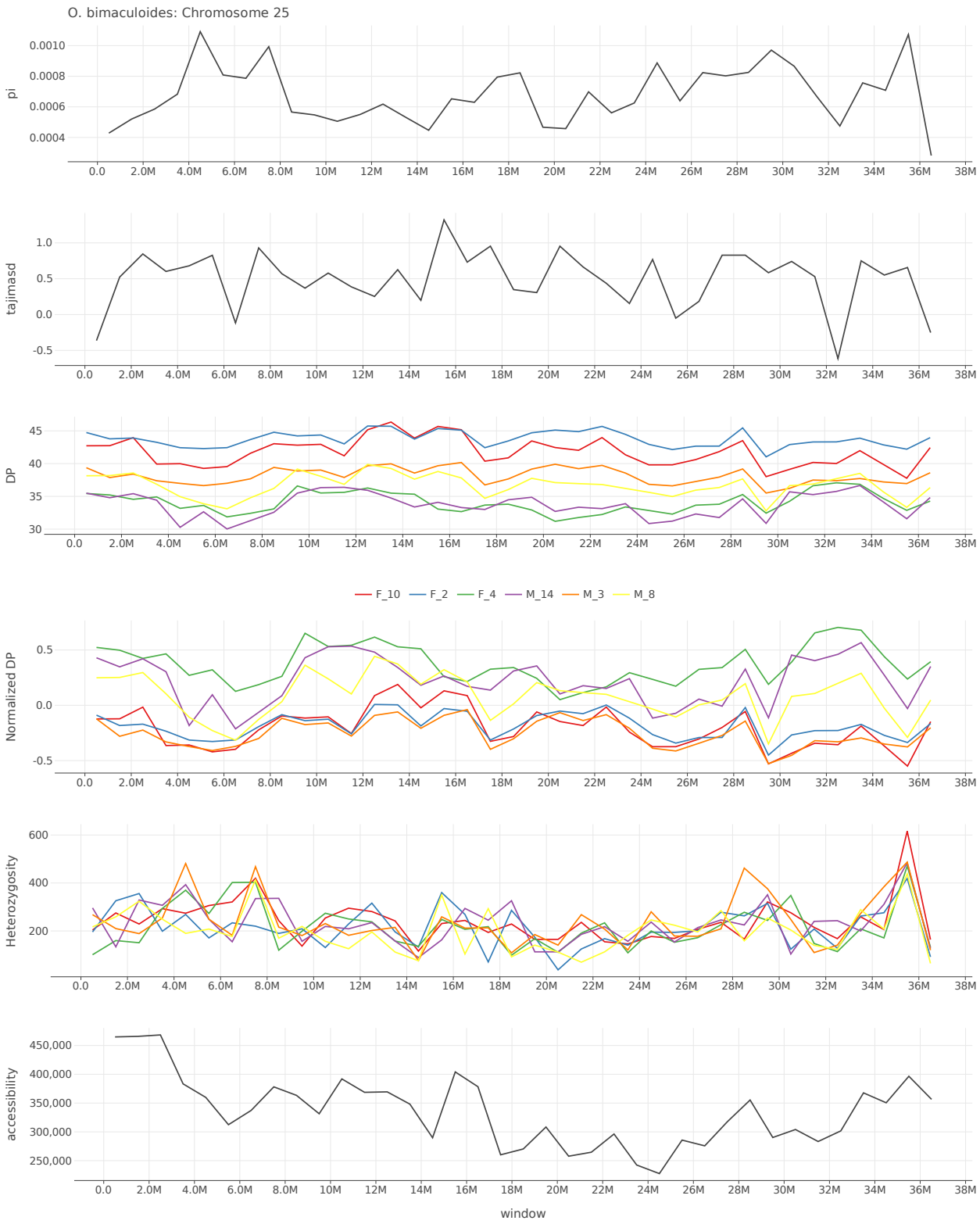

Figure S4.27: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 25 of *O. bimaculoides*.

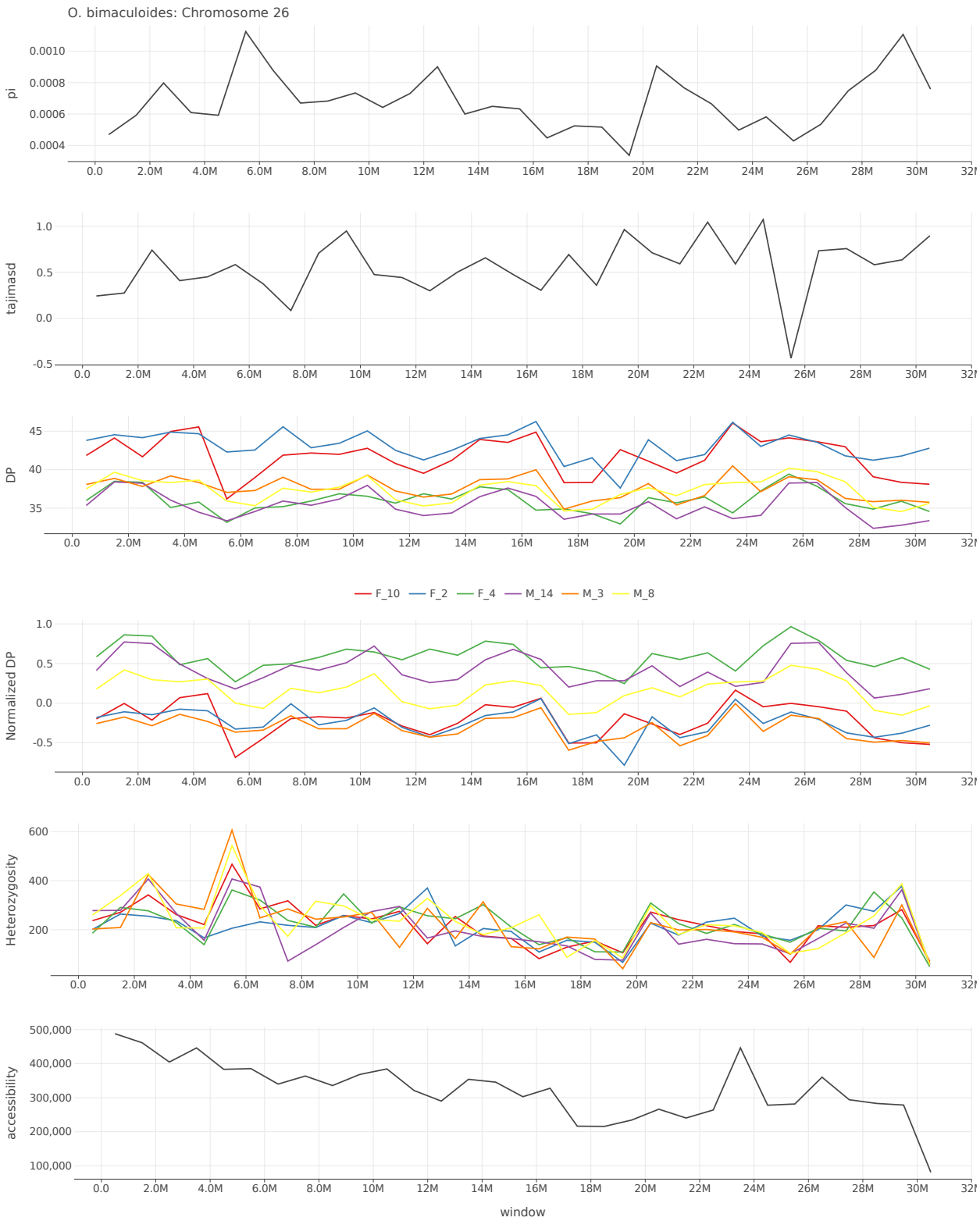

Figure S4.28: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 26 of *O. bimaculoides*.

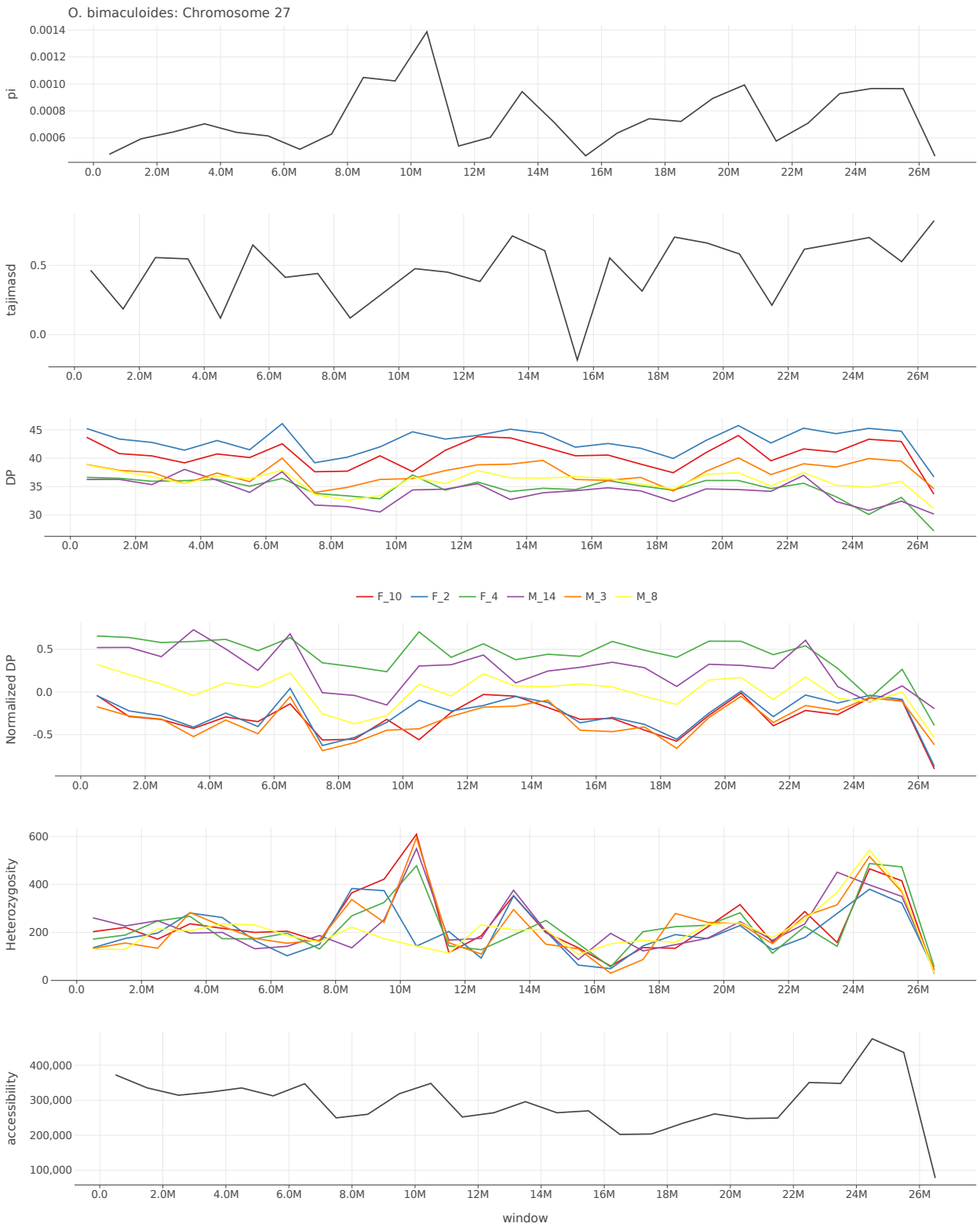

Figure S4.29: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 27 of *O. bimaculoides.*

Figure S4.30: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 28 of *O. bimaculoides*.

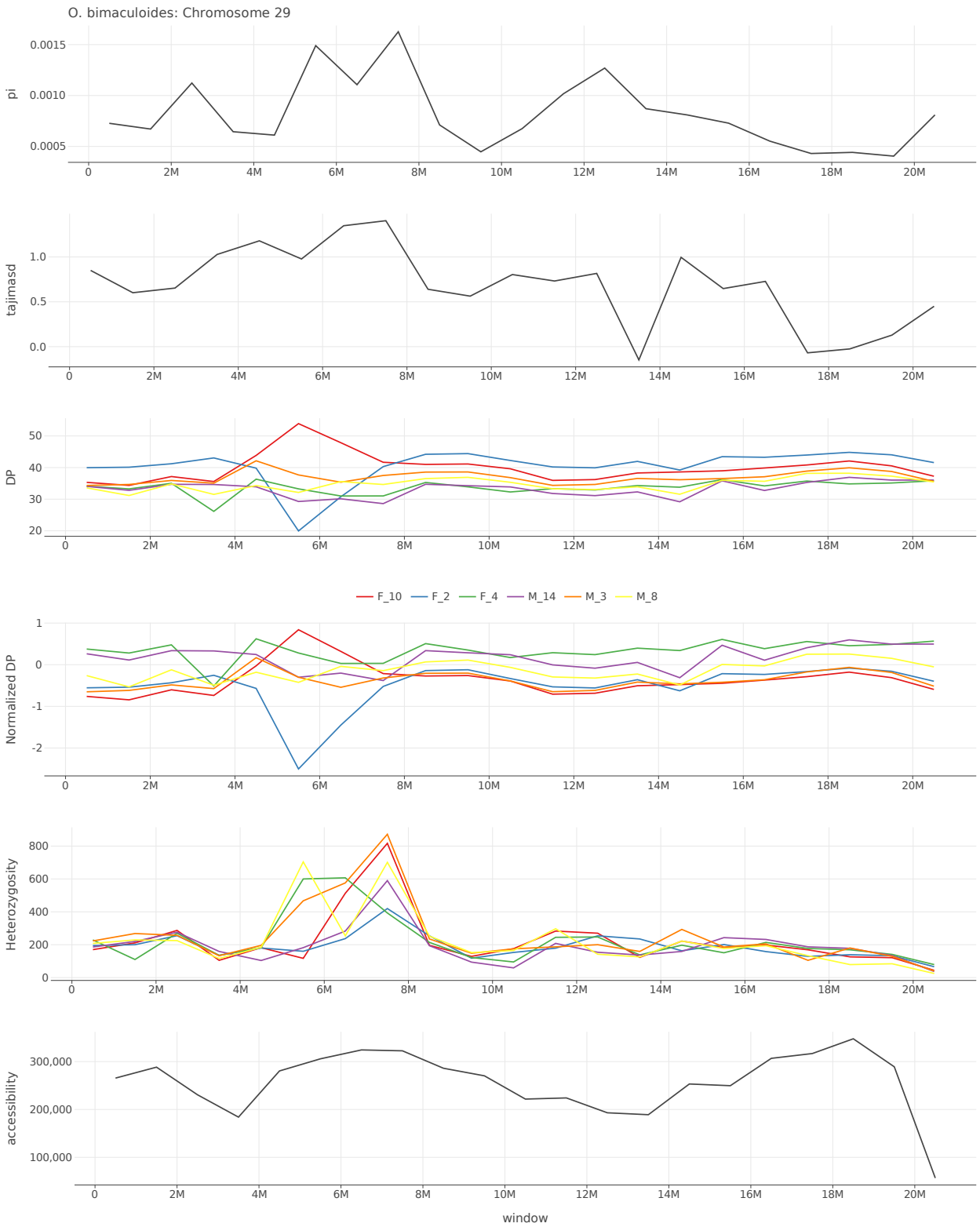

Figure S4.31: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 29 of *O. bimaculoides.*

Figure S4.32: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 30 of *O. bimaculoides.*

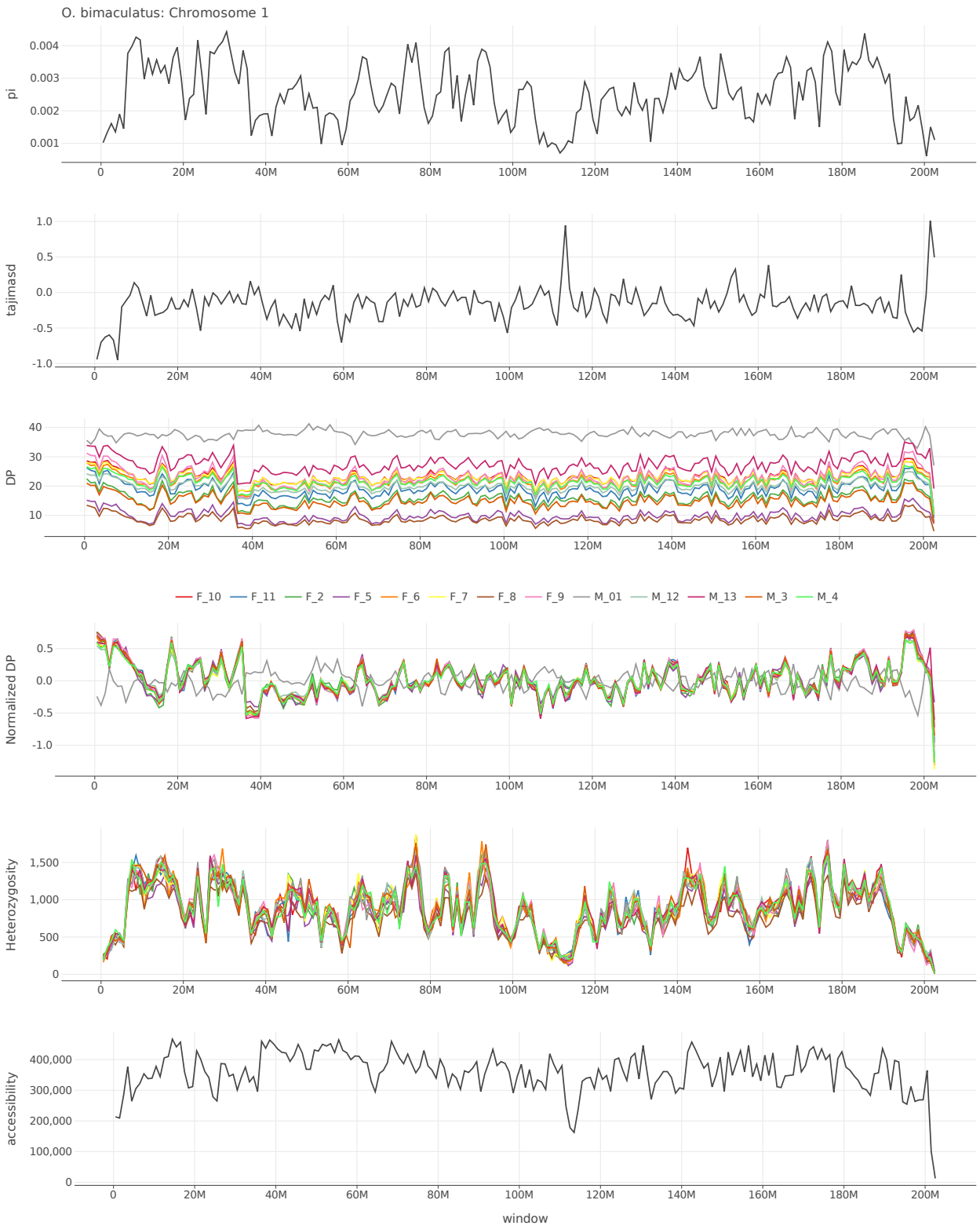

Figure S4.33: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 1 of *O. bimaculatus*.

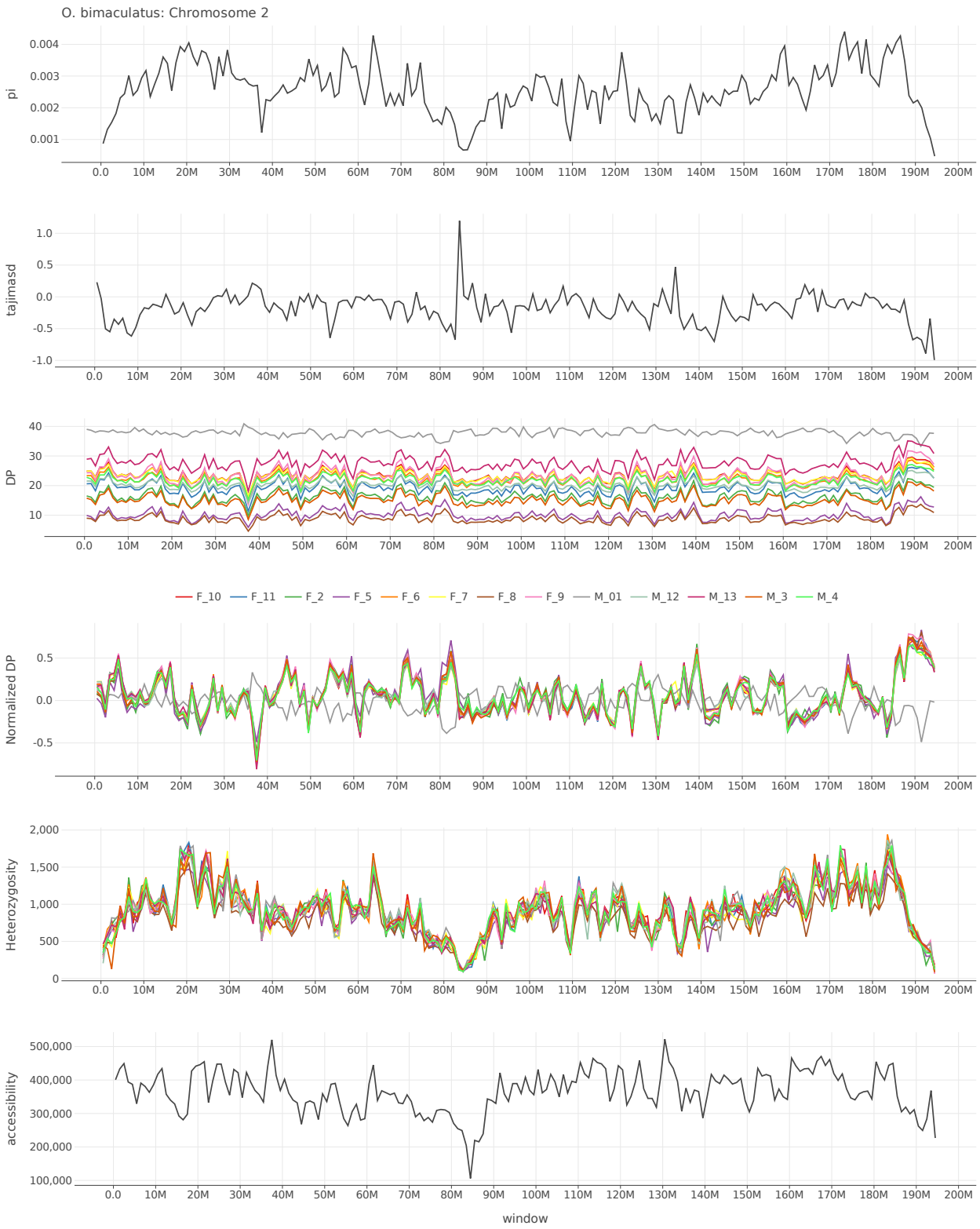

Figure S4.34: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 2 of *O. bimaculatus*.

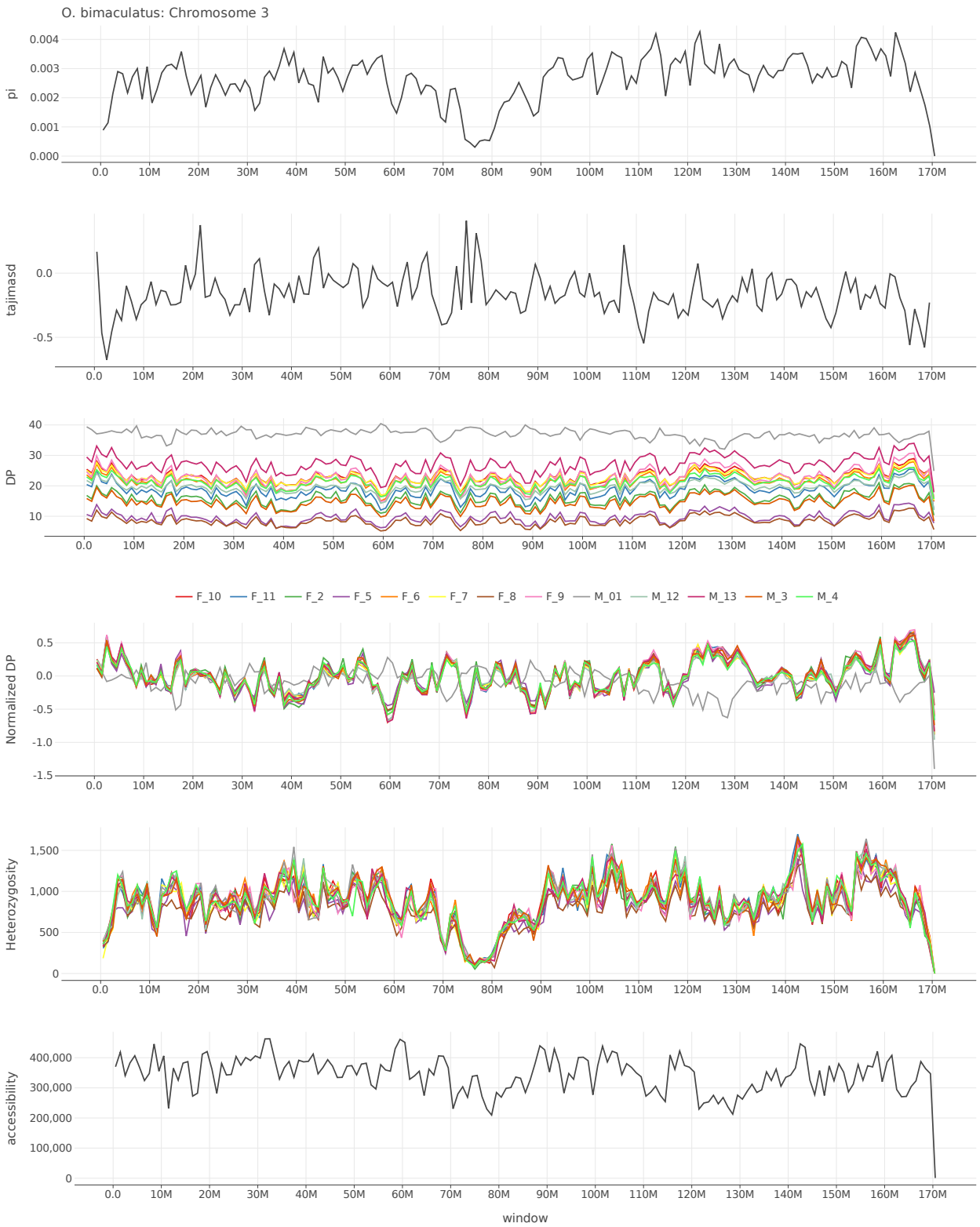

Figure S4.35: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 3 of *O. bimaculatus*.

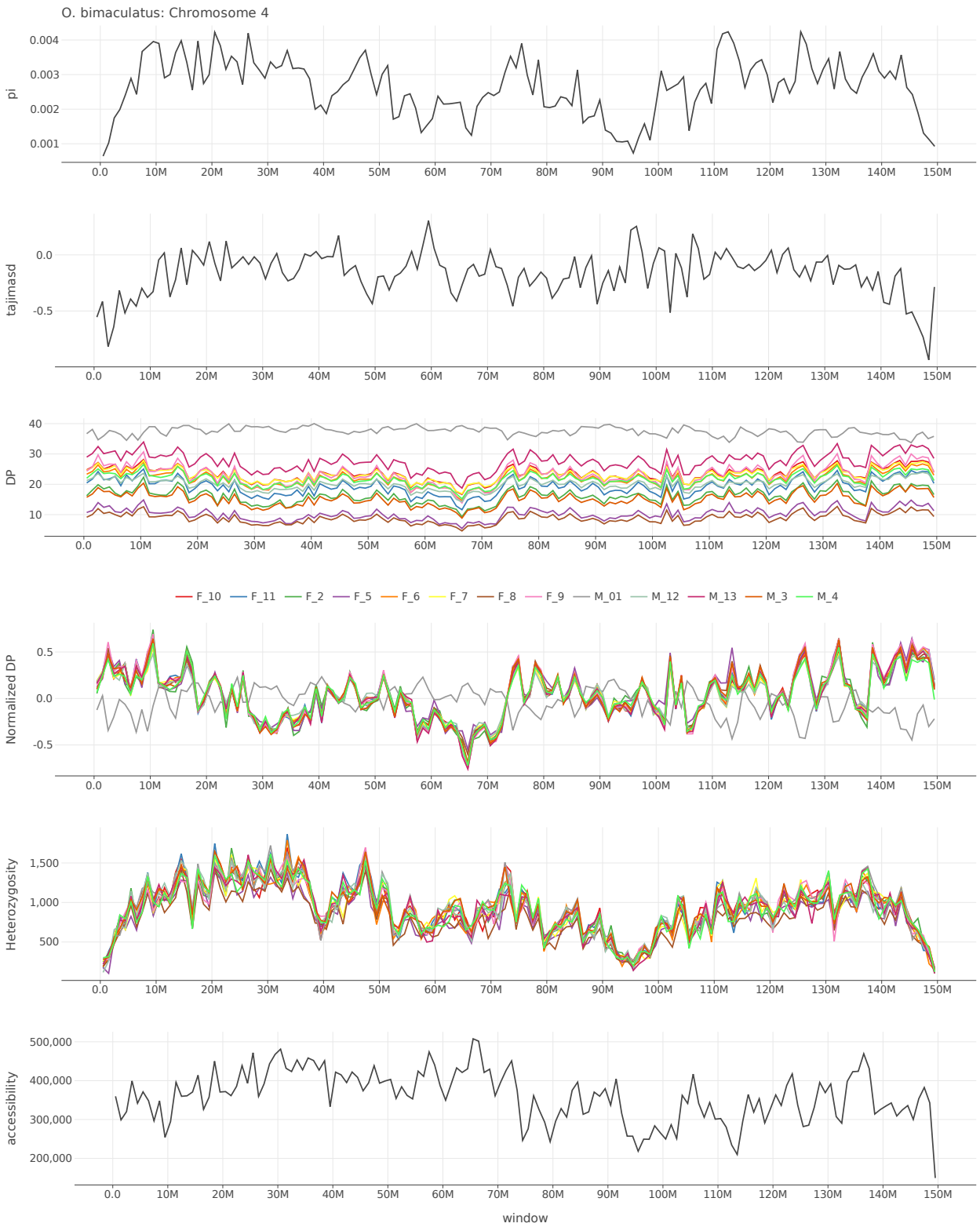

Figure S4.36: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 4 of *O. bimaculatus*.

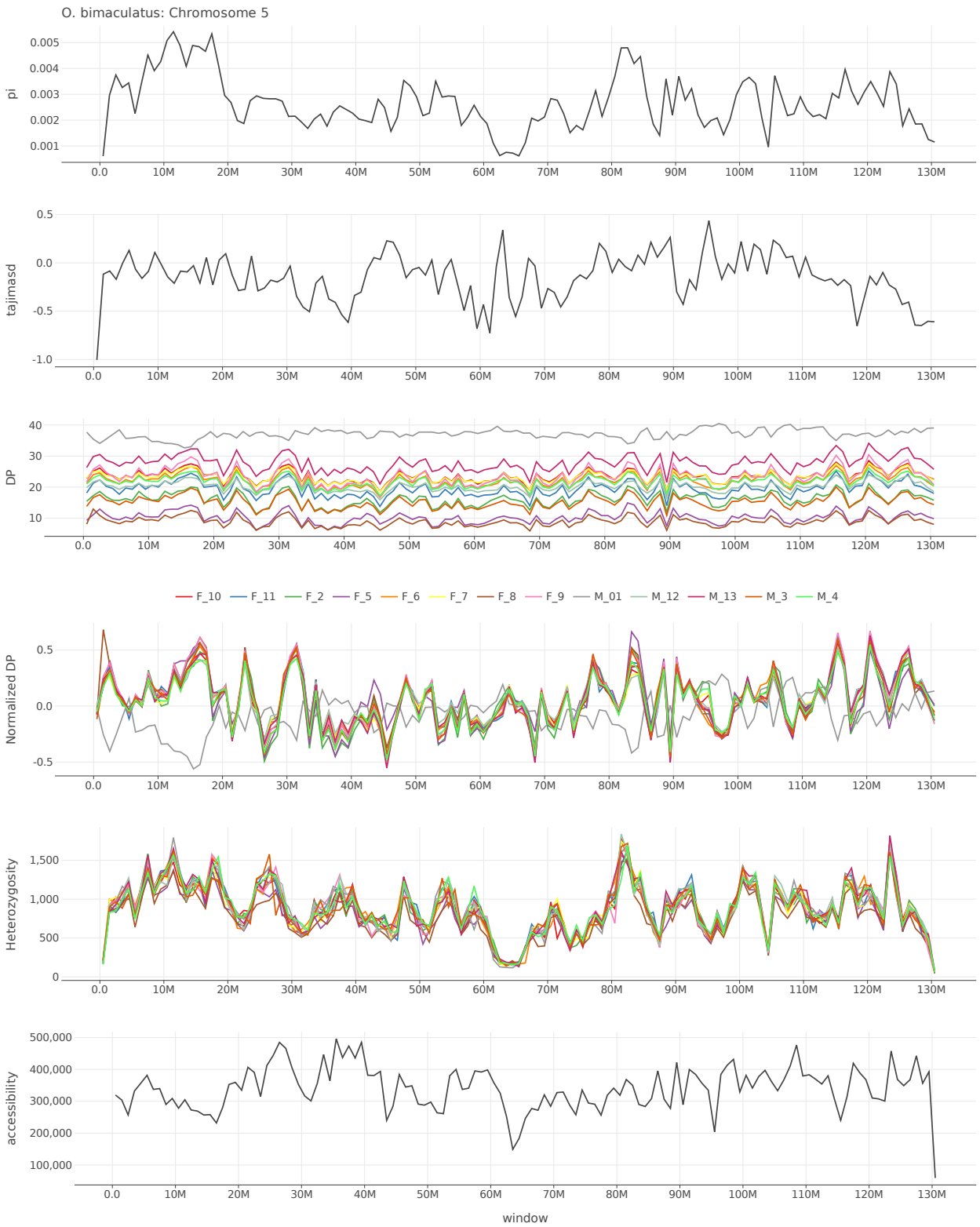

Figure S4.37: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 5 of *O. bimaculatus*.

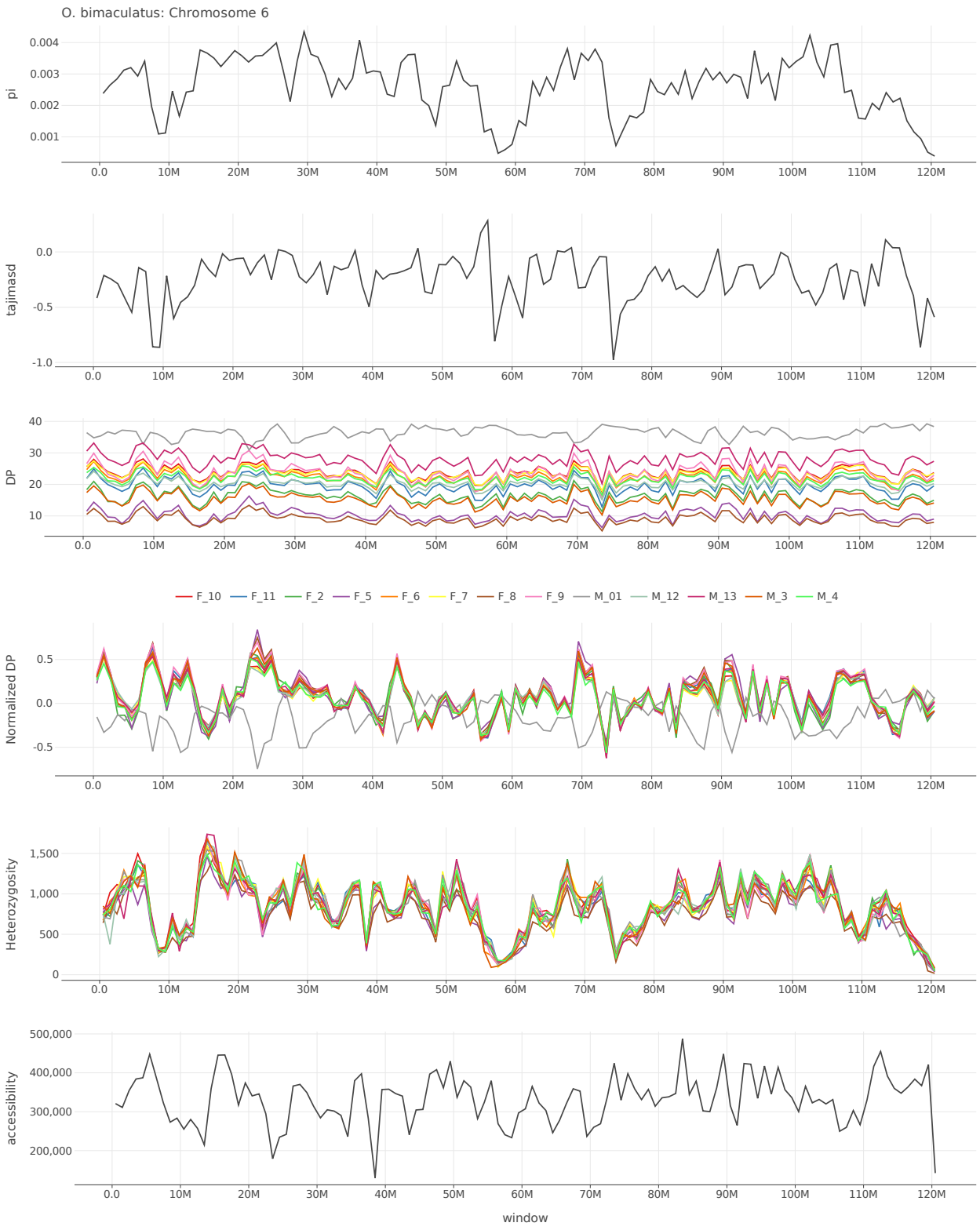

Figure S4.38: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 6 of *O. bimaculatus.*

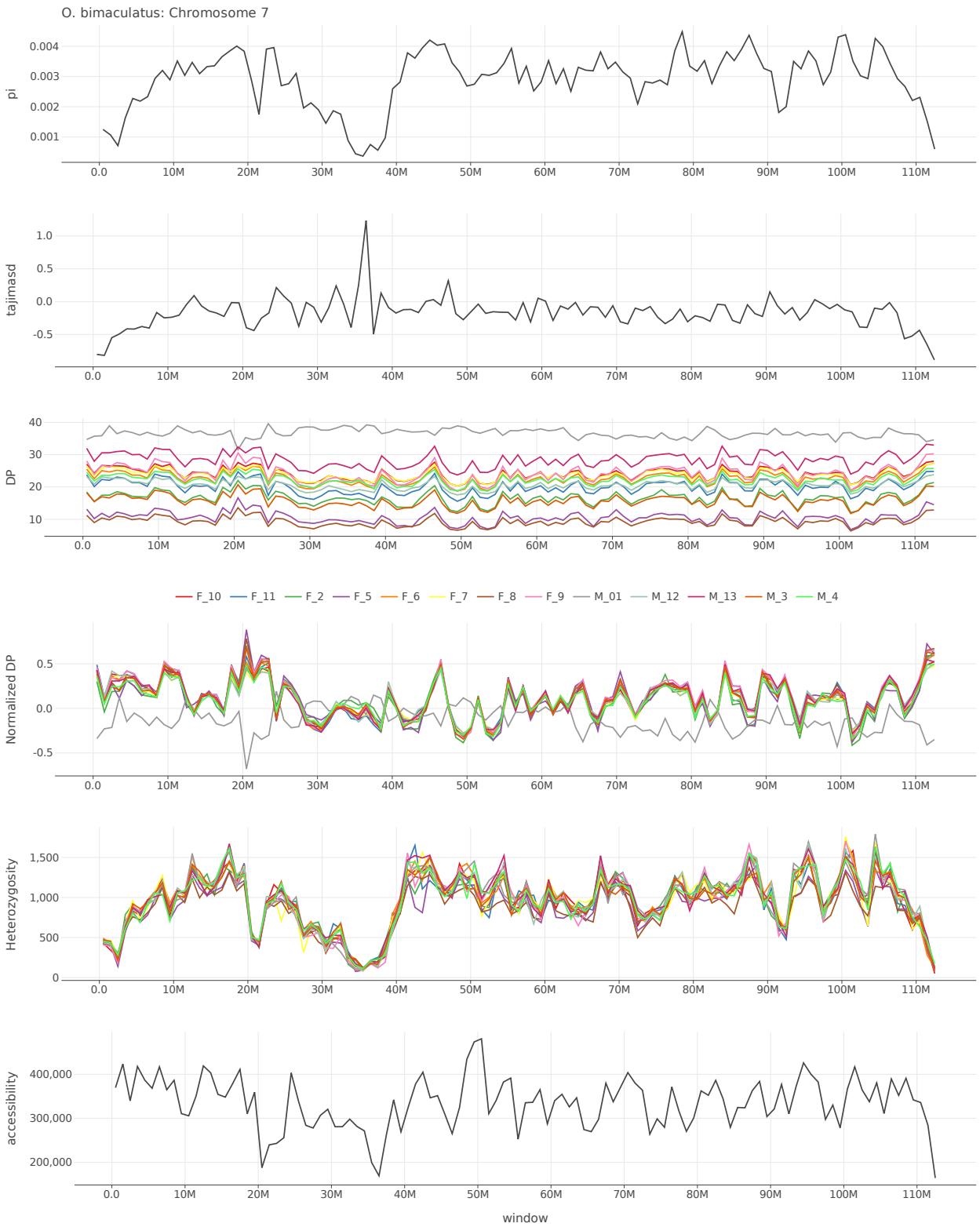

Figure S4.39: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 7 of *O. bimaculatus*.

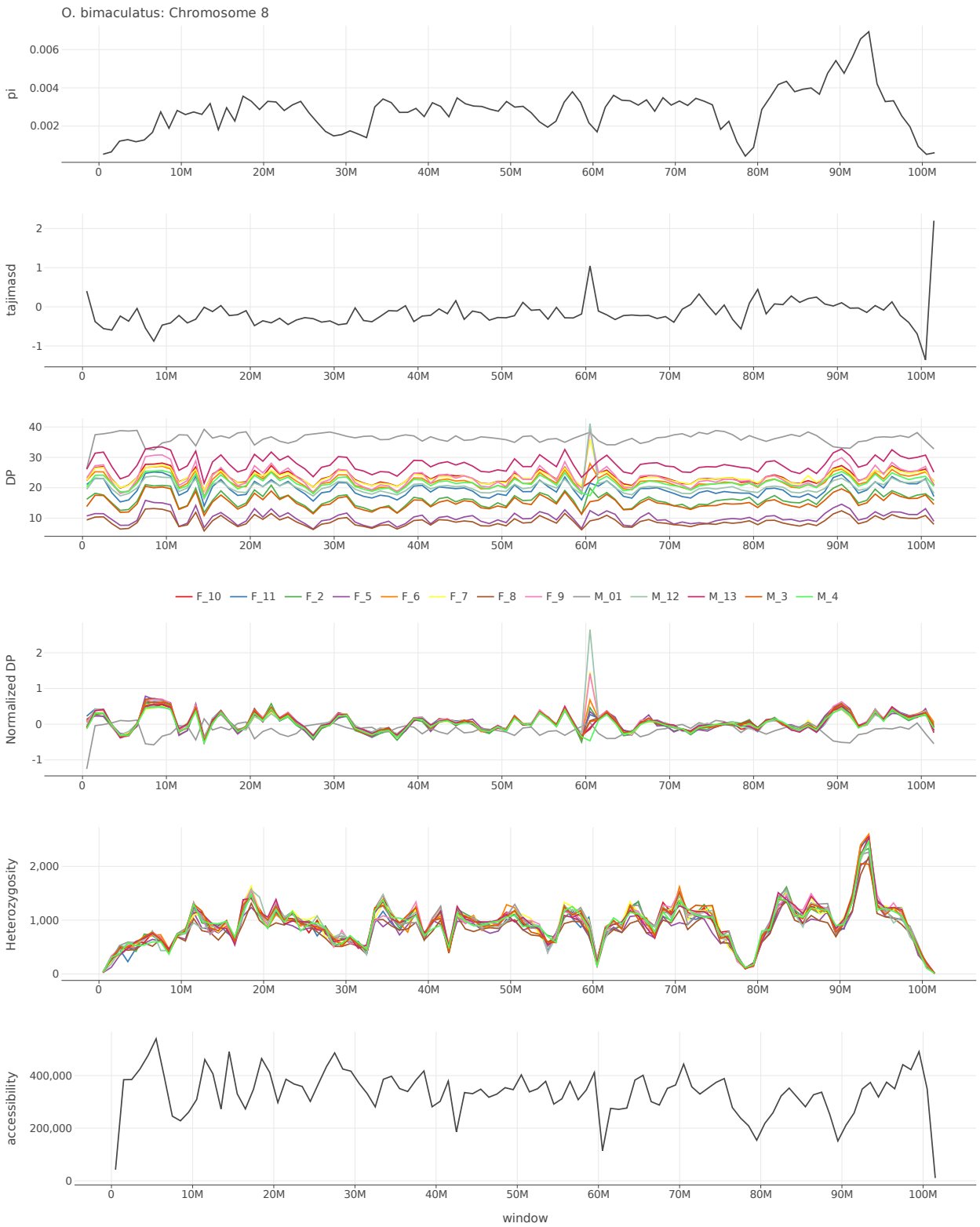

Figure S4.40: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 8 of *O. bimaculatus*.

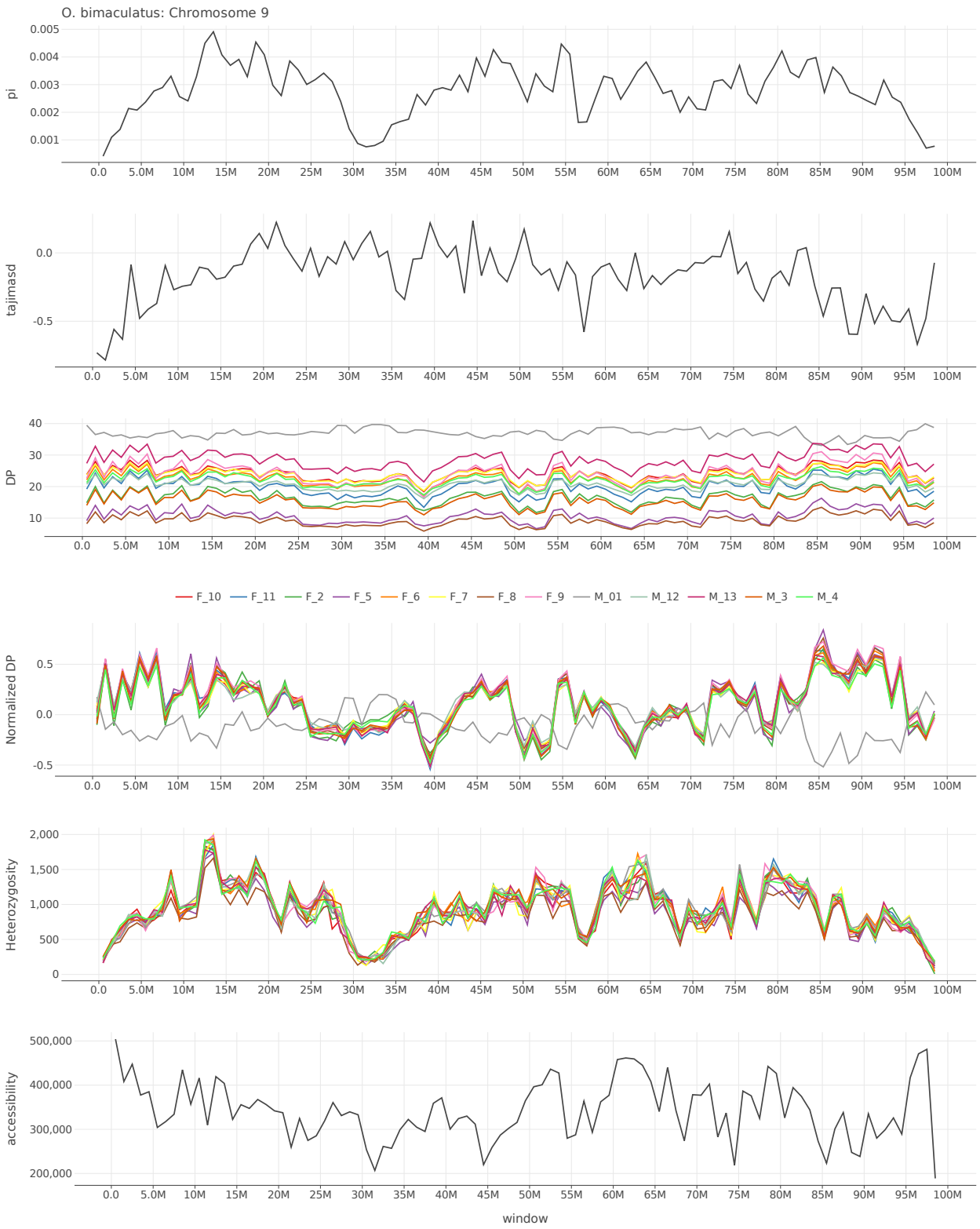

Figure S4.41: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 9 of *O. bimaculatus*.

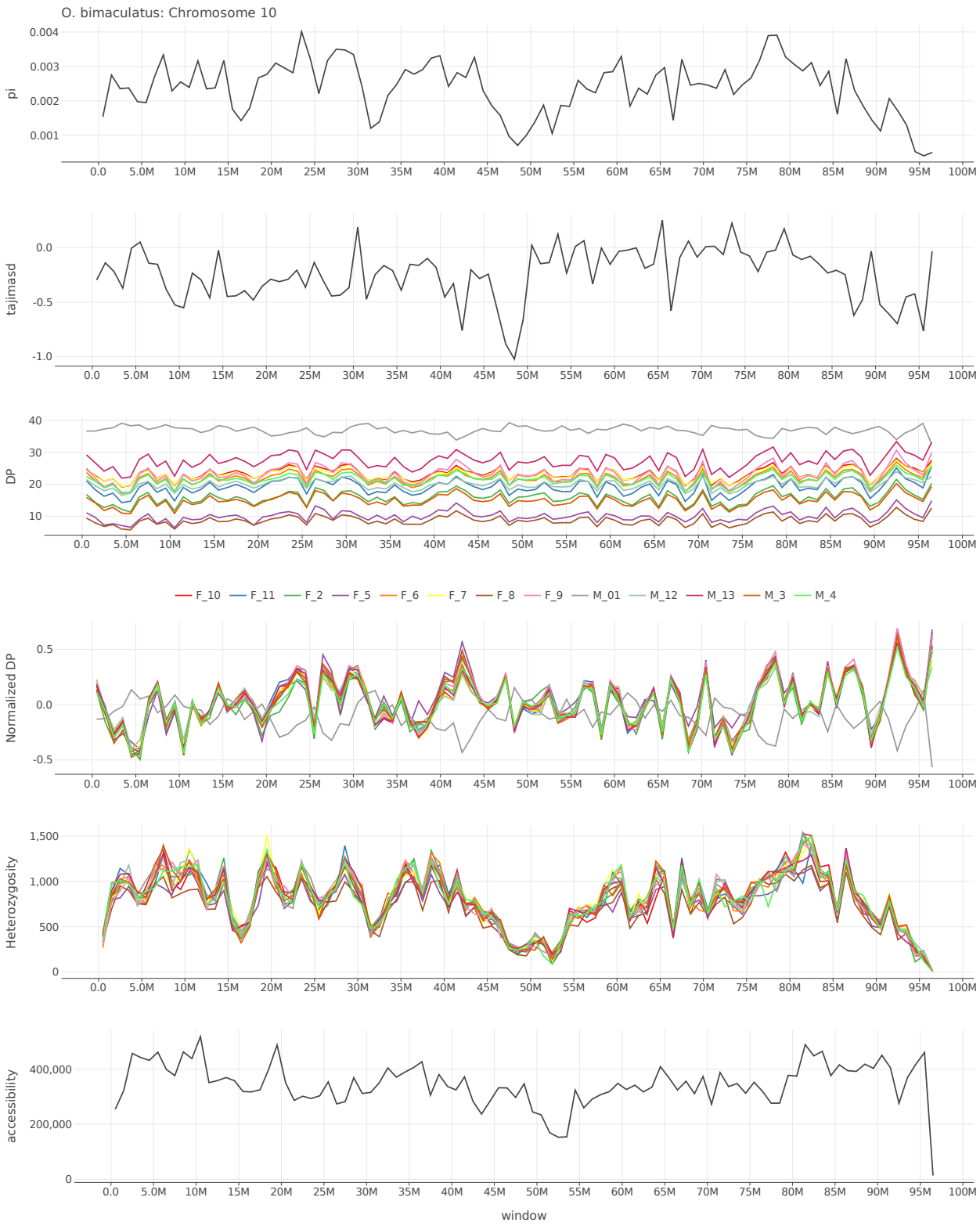

Figure S4.42: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 10 of *O. bimaculatus.*

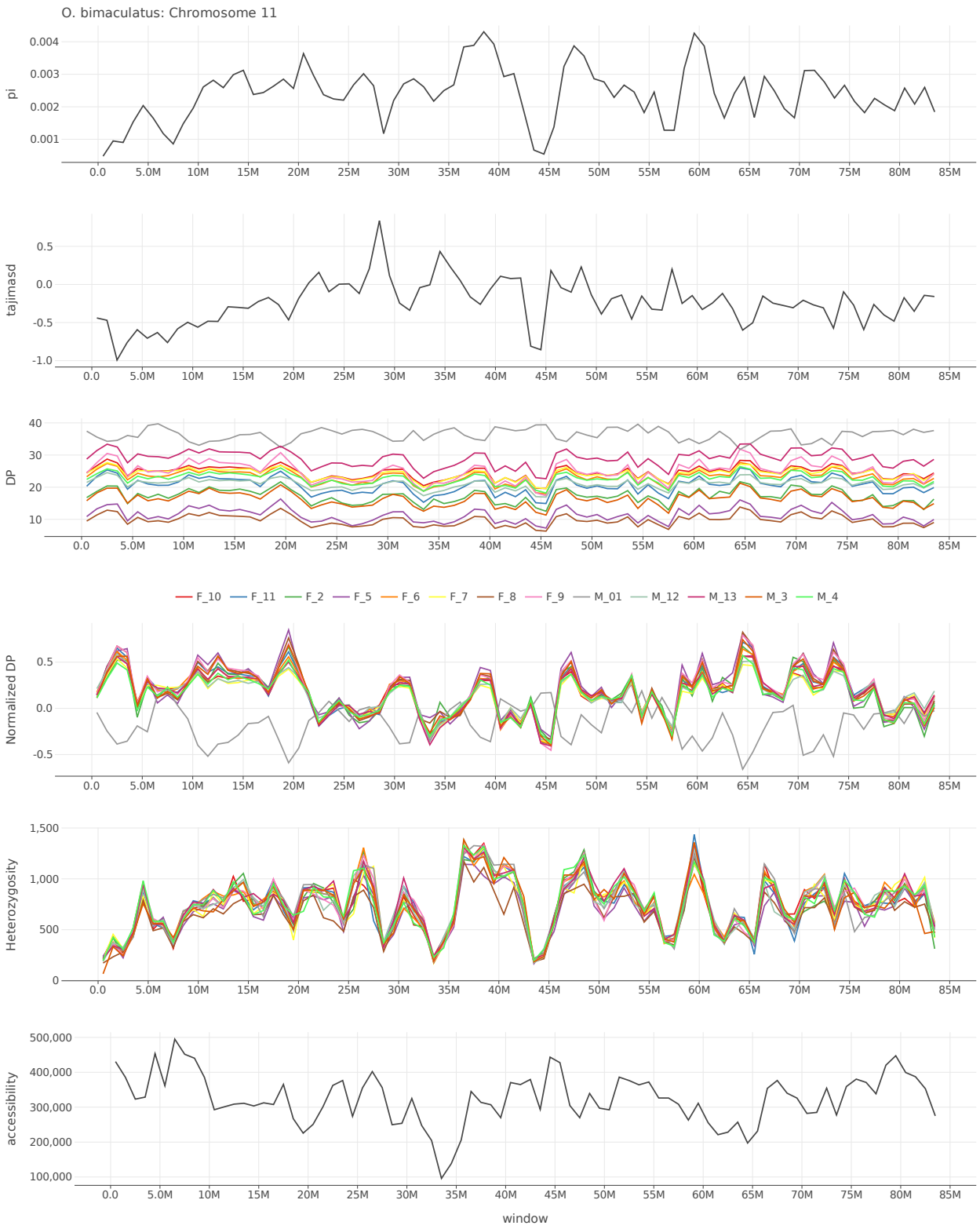

Figure S4.43: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 11 of *O. bimaculatus.*

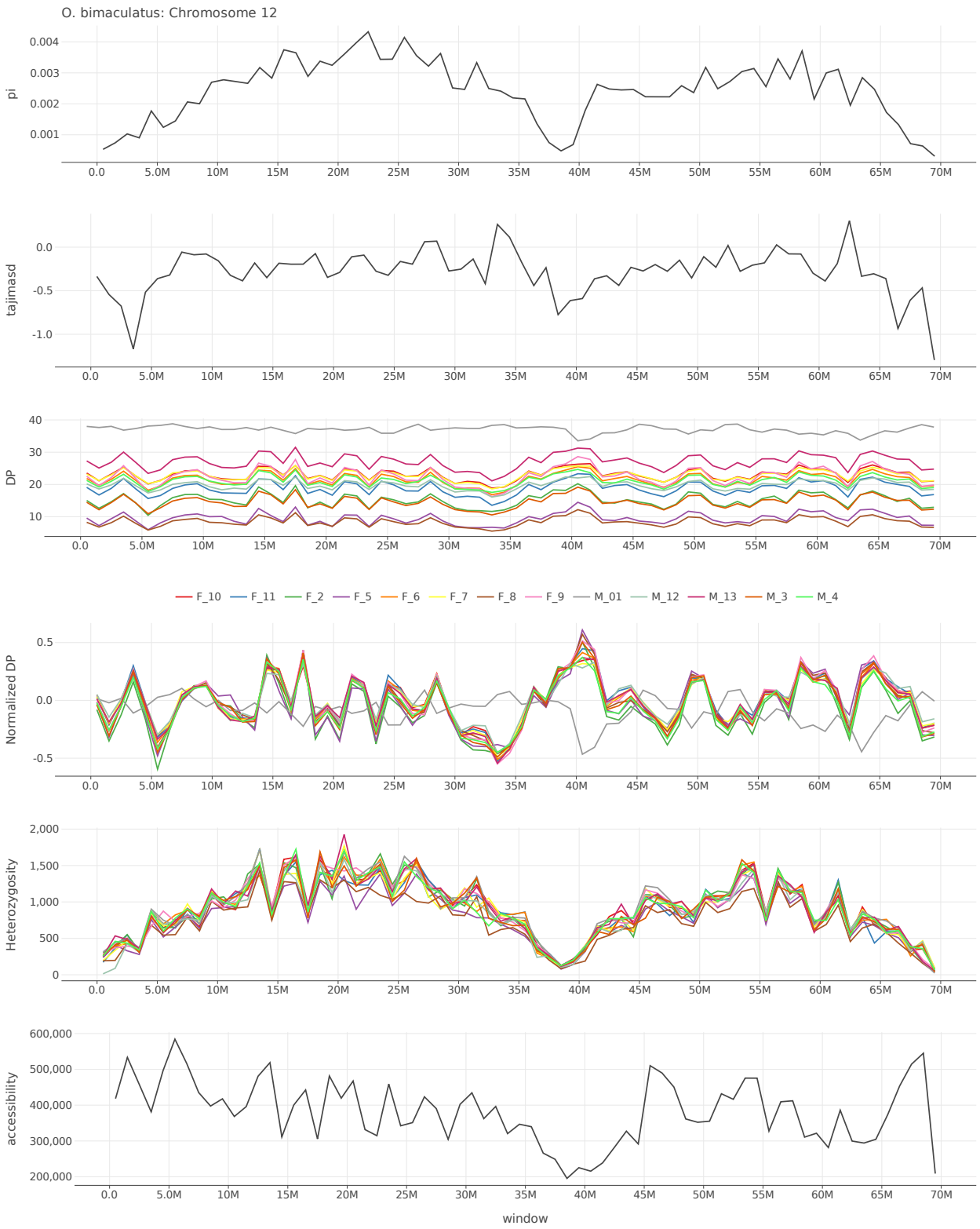

Figure S4.4:4 Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 12 of *O. bimaculatus.*

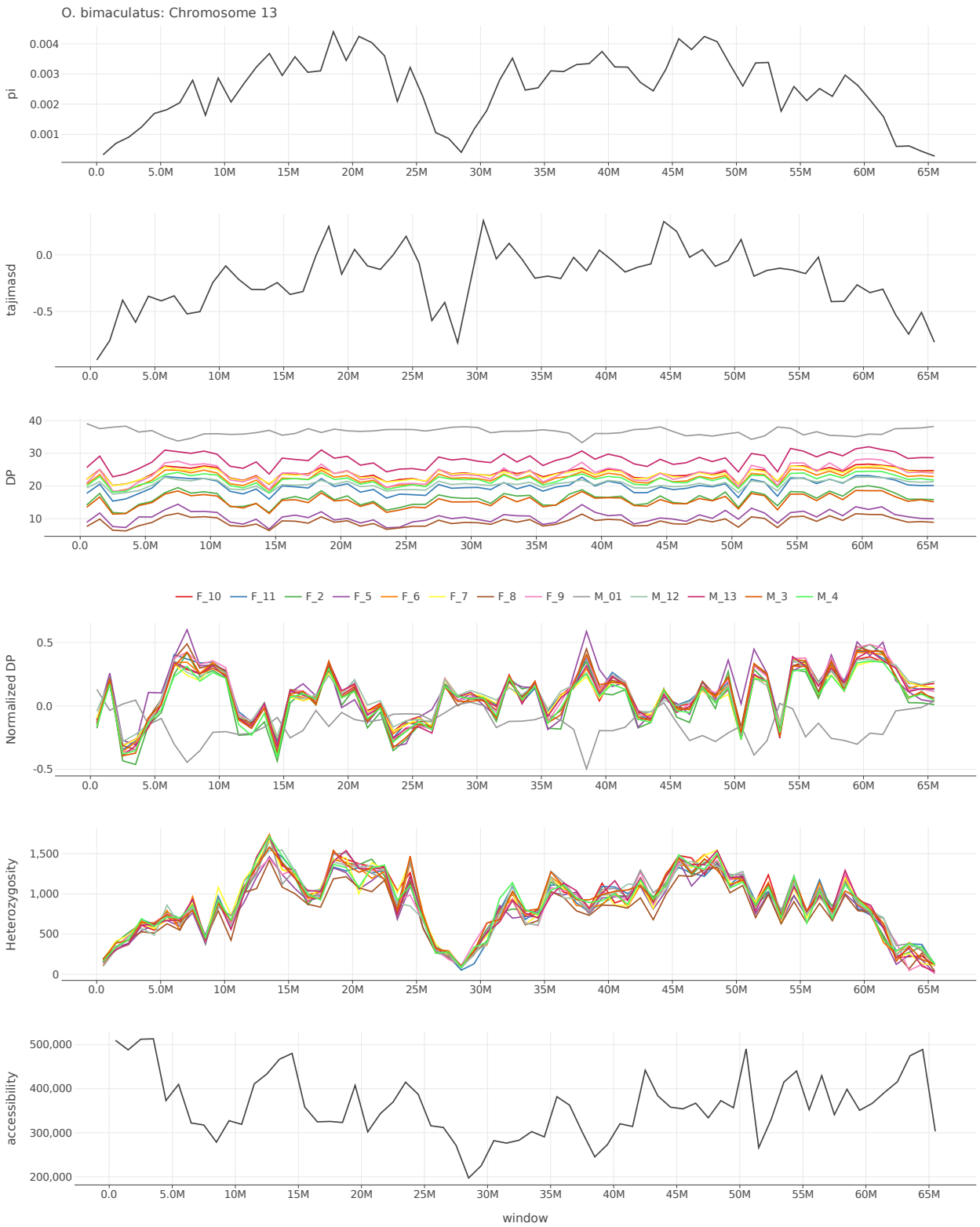

Figure S4.45: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 13 of *O. bimaculatus.*

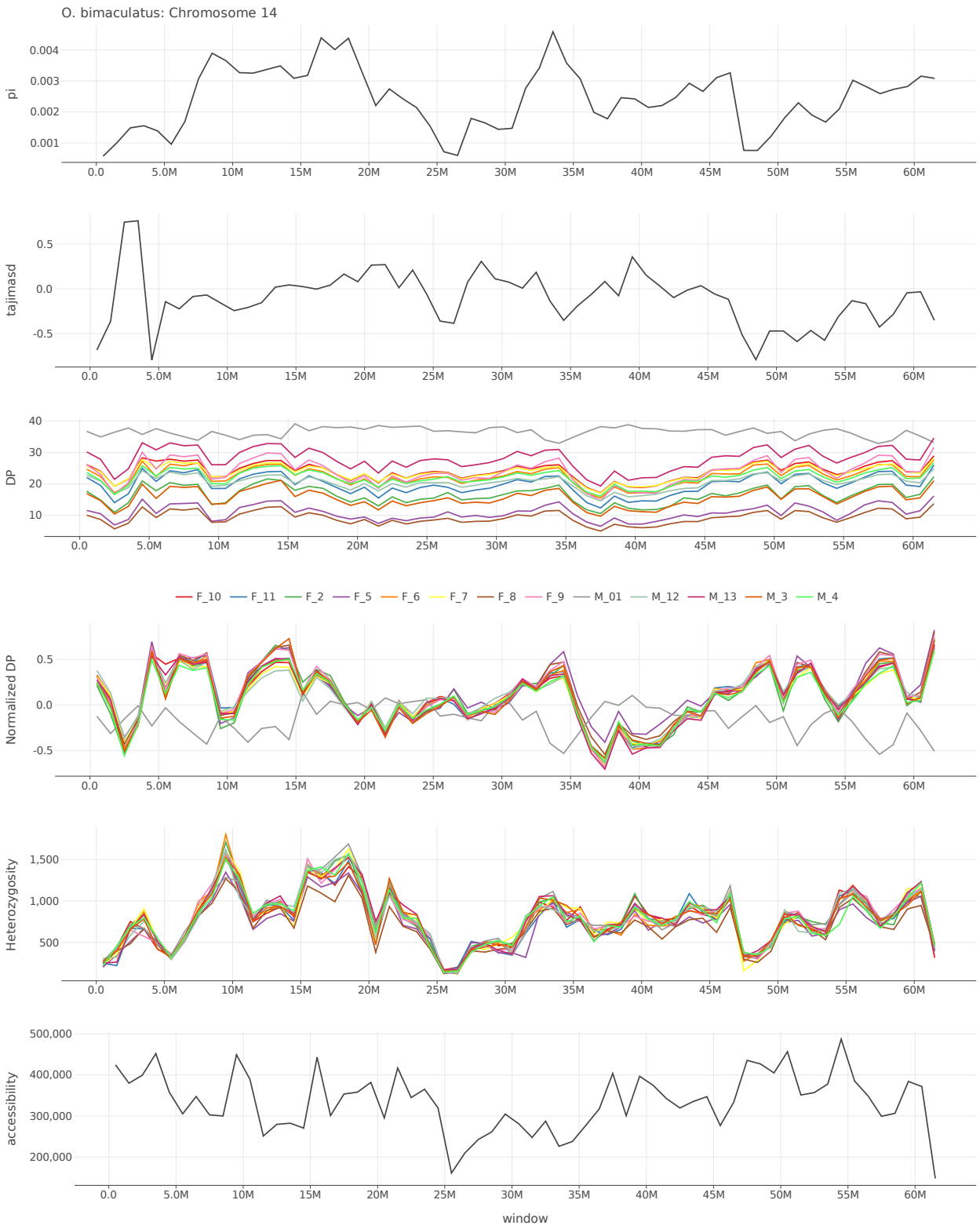

Figure S4.46: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 14 of *O. bimaculatus.*

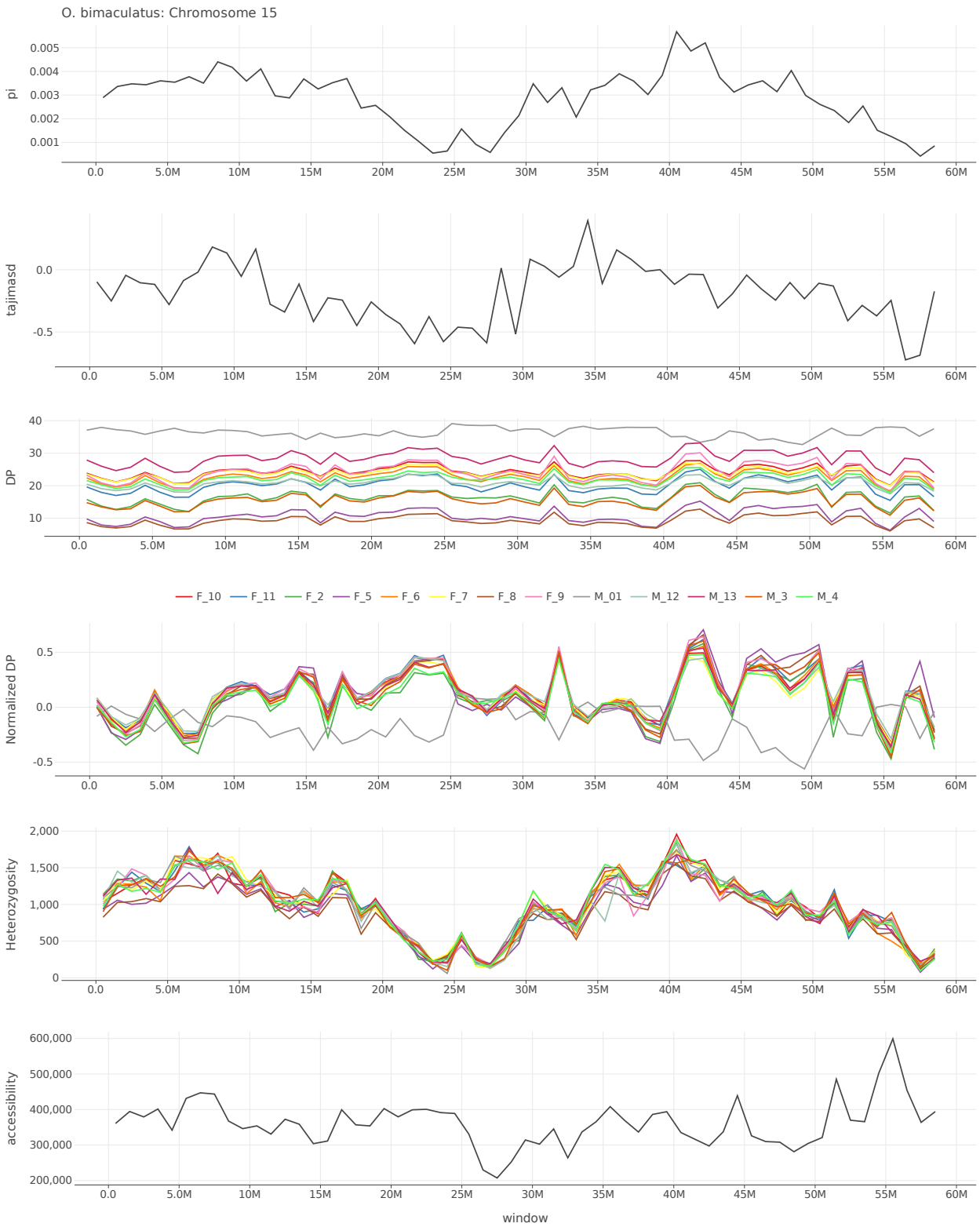

Figure S4.47: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 15 of *O. bimaculatus.*

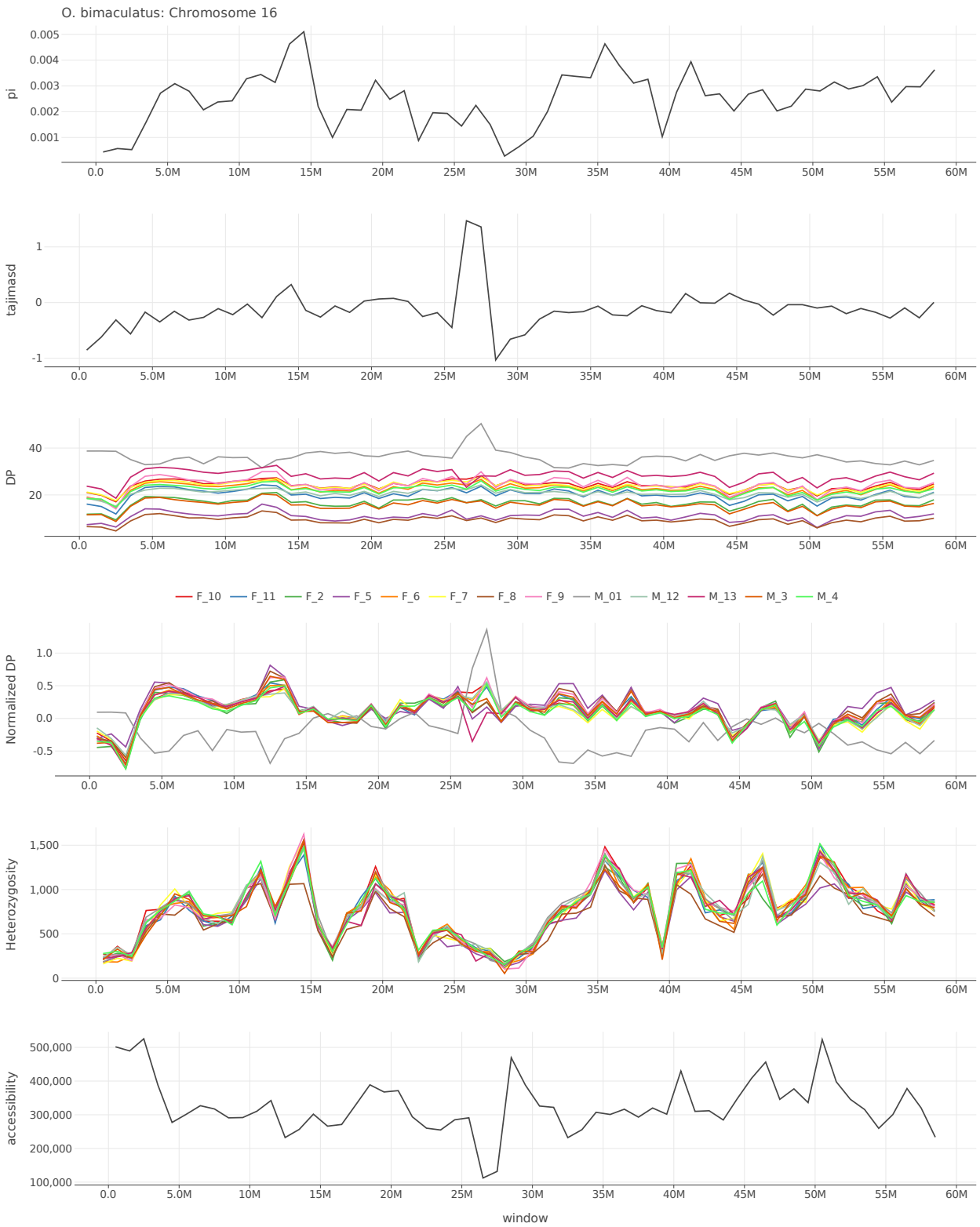

Figure S4.48: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 16 of *O. bimaculatus.*

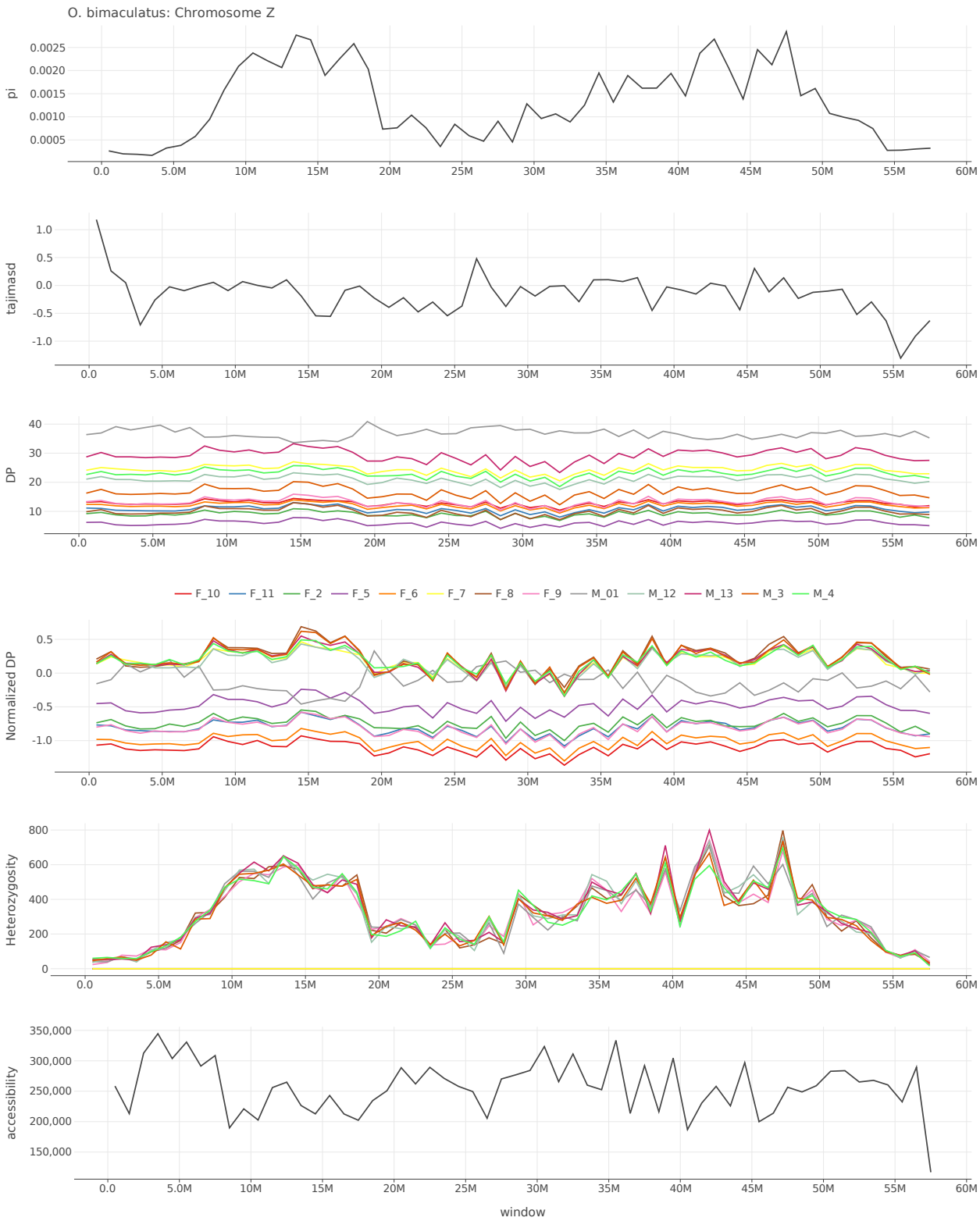

Figure S4.49: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome Z of *O. bimaculatus.*

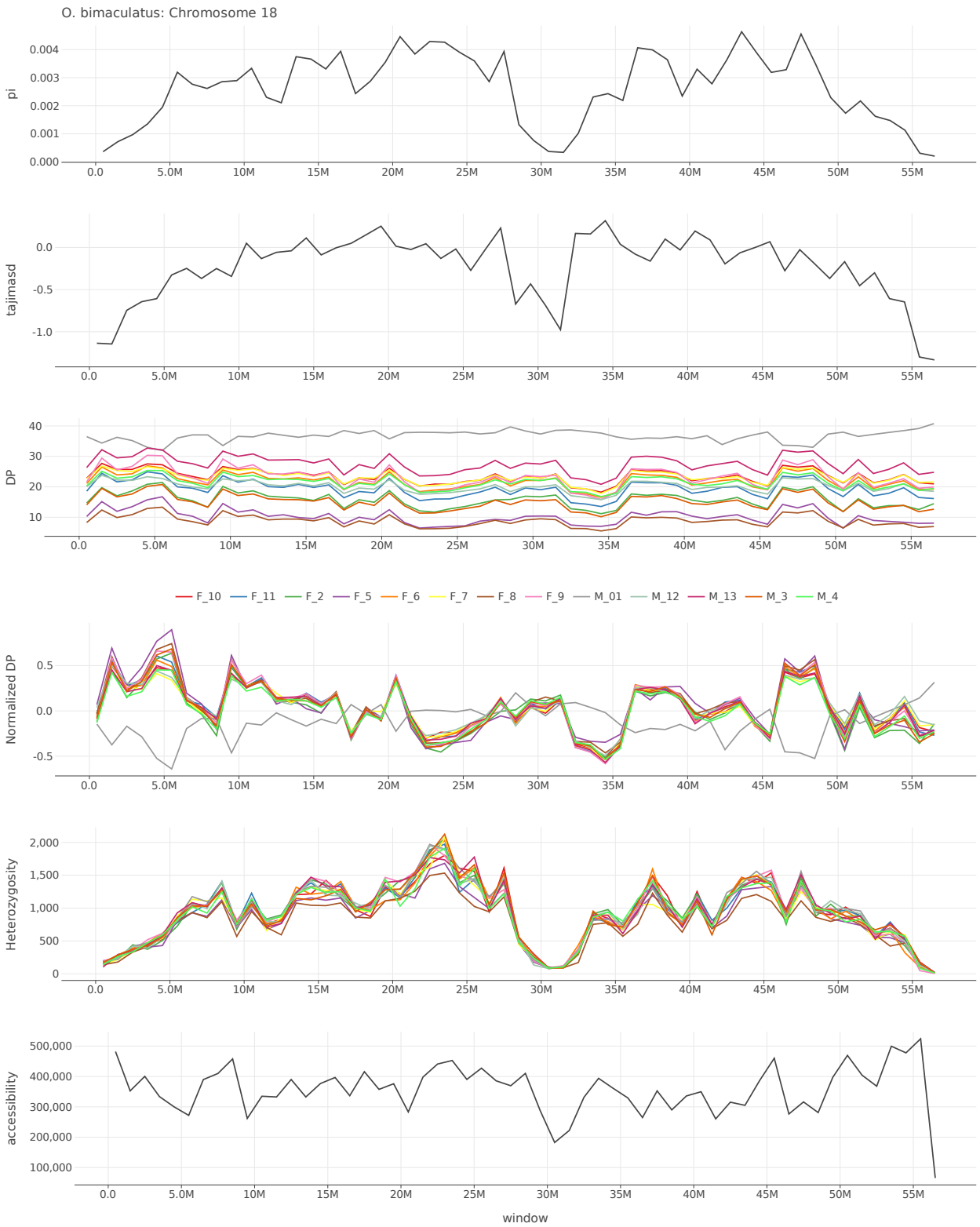

Figure S4.50: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 18 of *O. bimaculatus.*

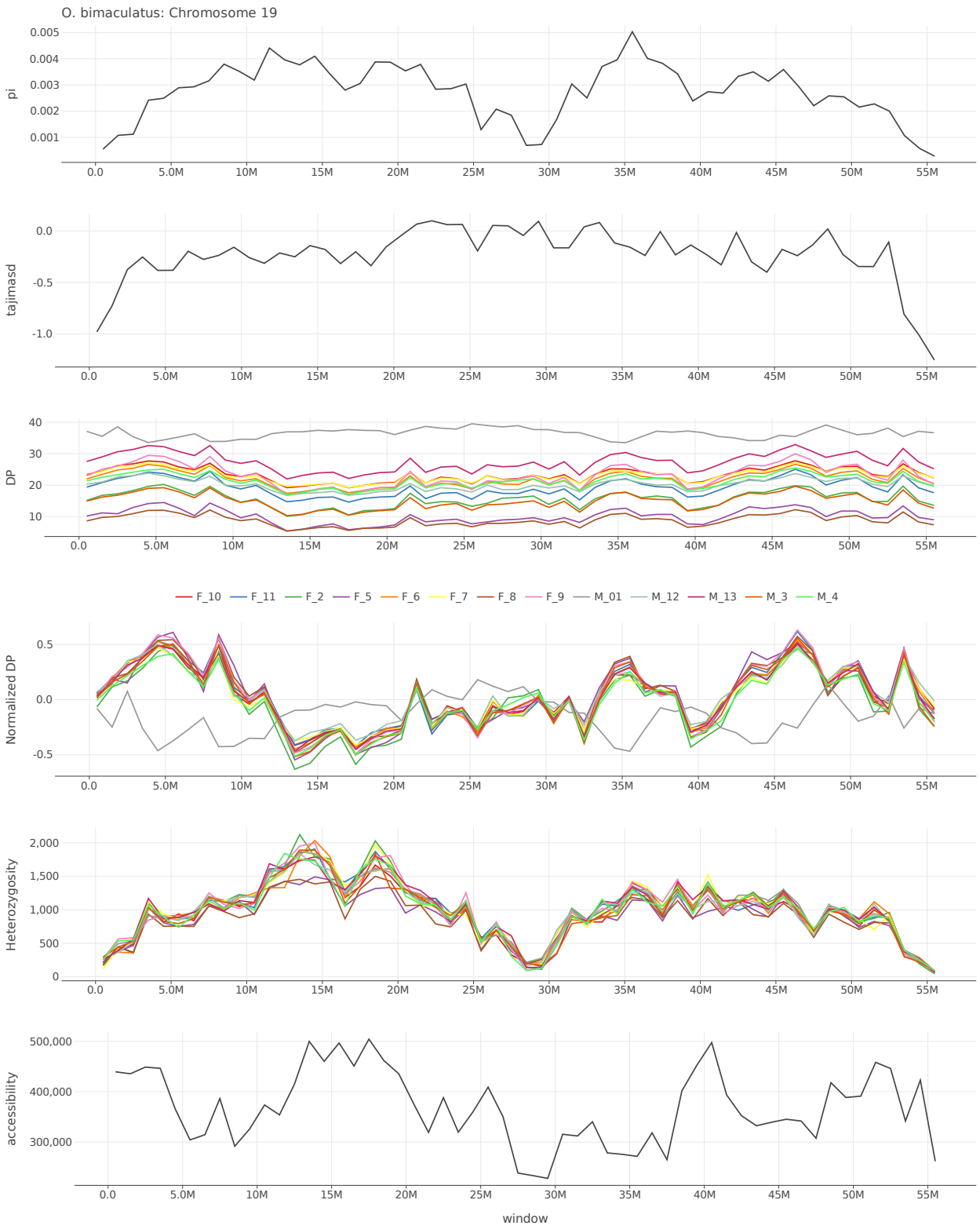

Figure S4.51: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 19 of *O. bimaculatus.*

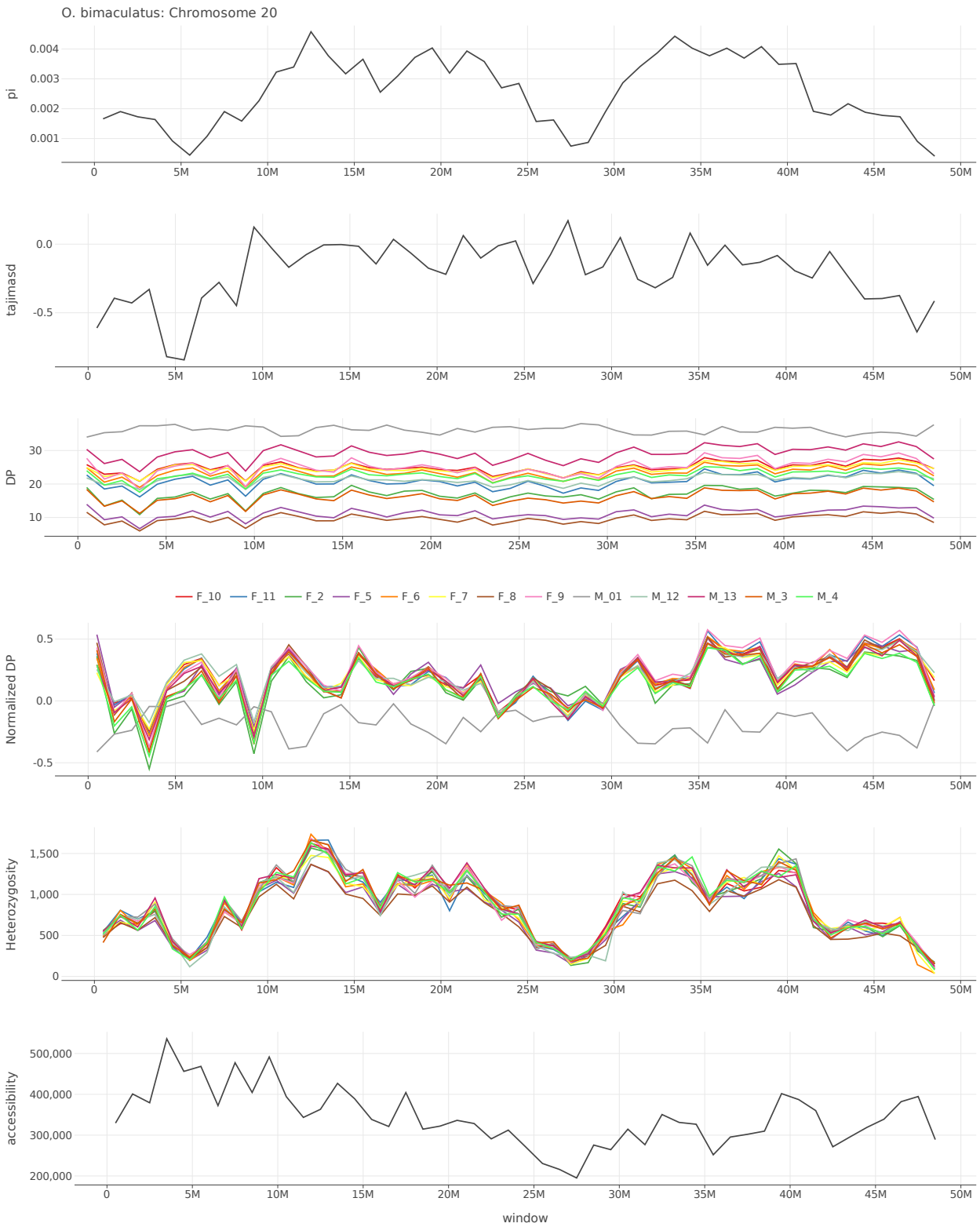

Figure S4.52: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 20 of *O. bimaculatus.*

Figure S4.53: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 21 of *O. bimaculatus.*

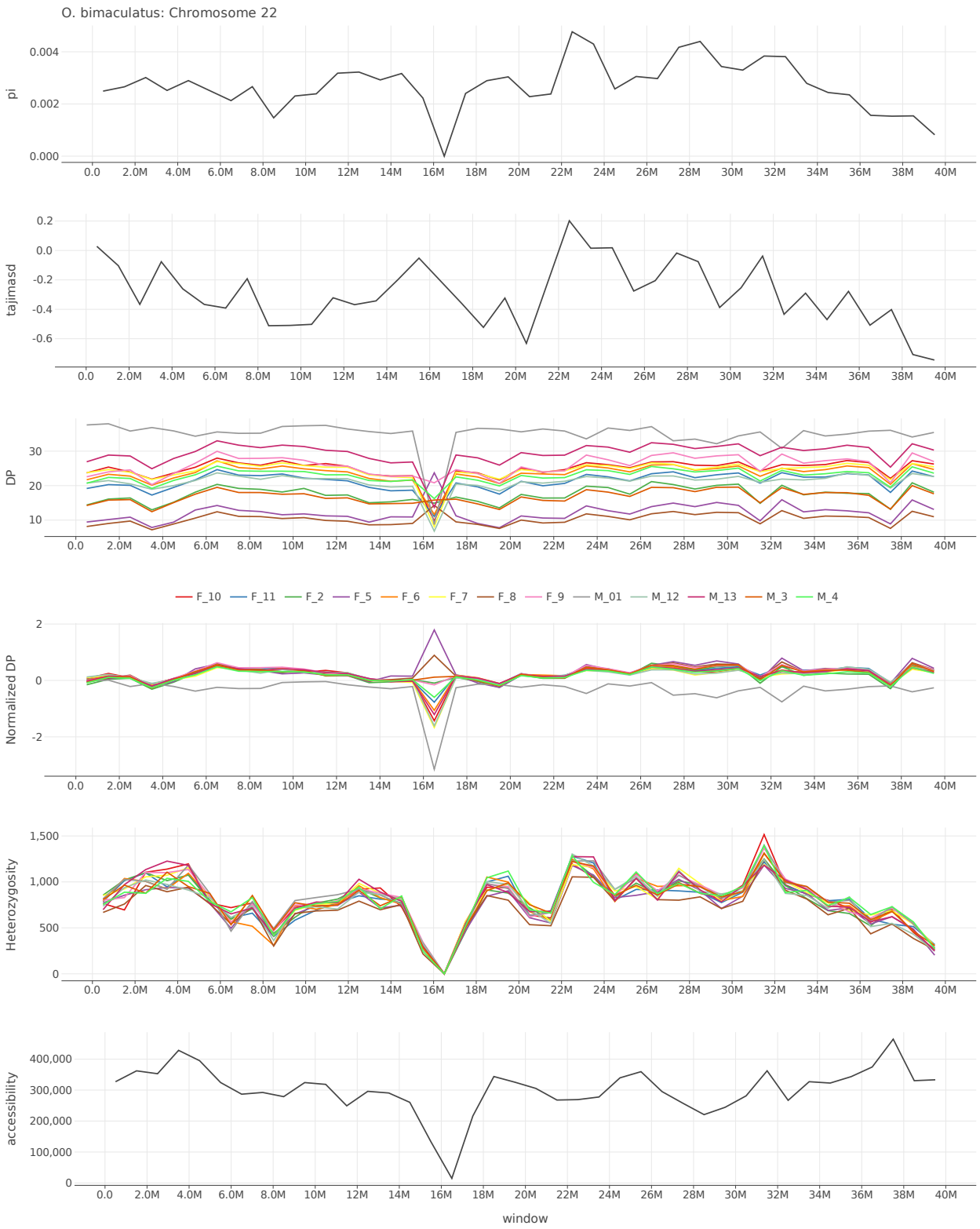

Figure S4.54: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 22 of *O. bimaculatus.*

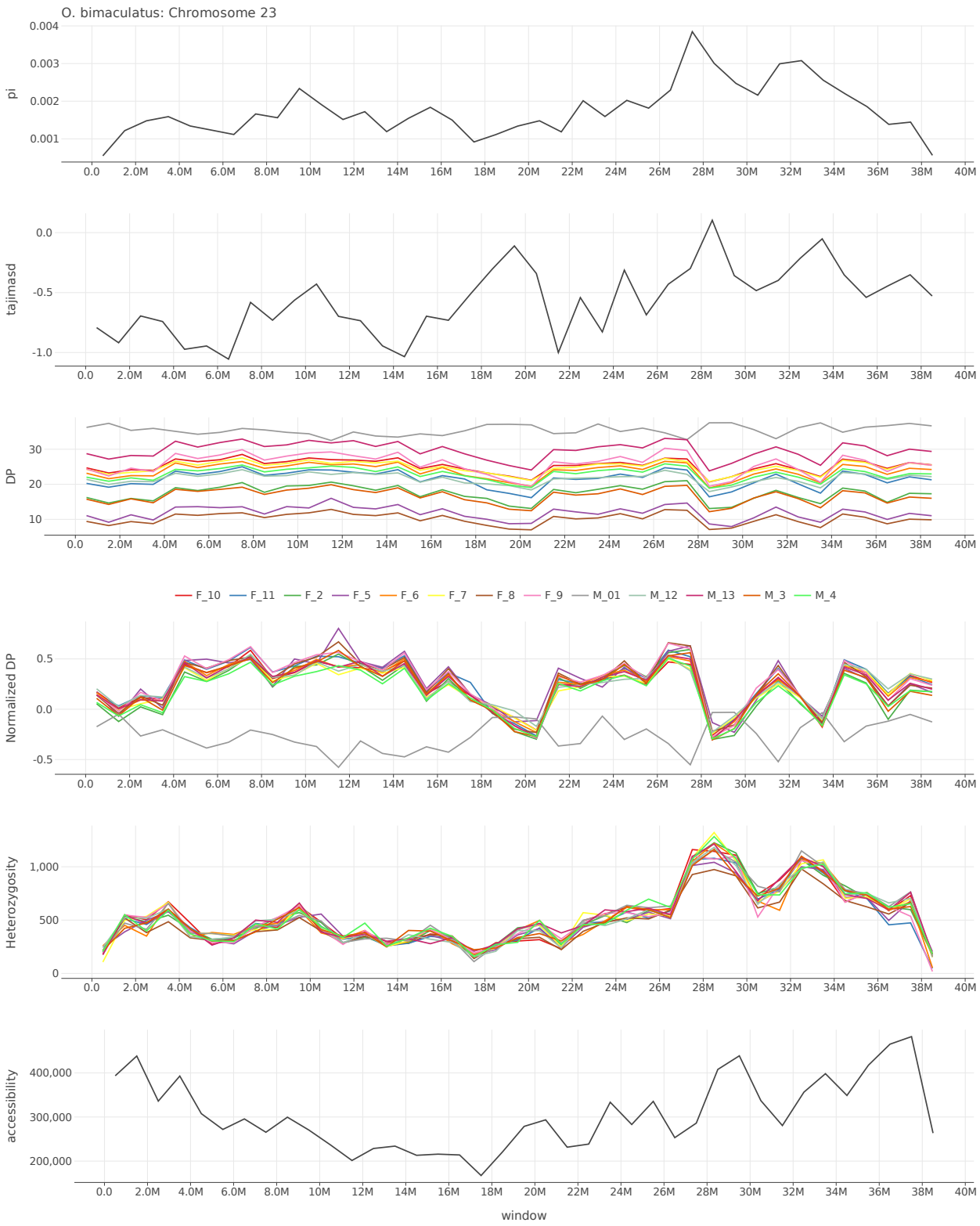

Figure S4.55: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 23 of *O. bimaculatus.*

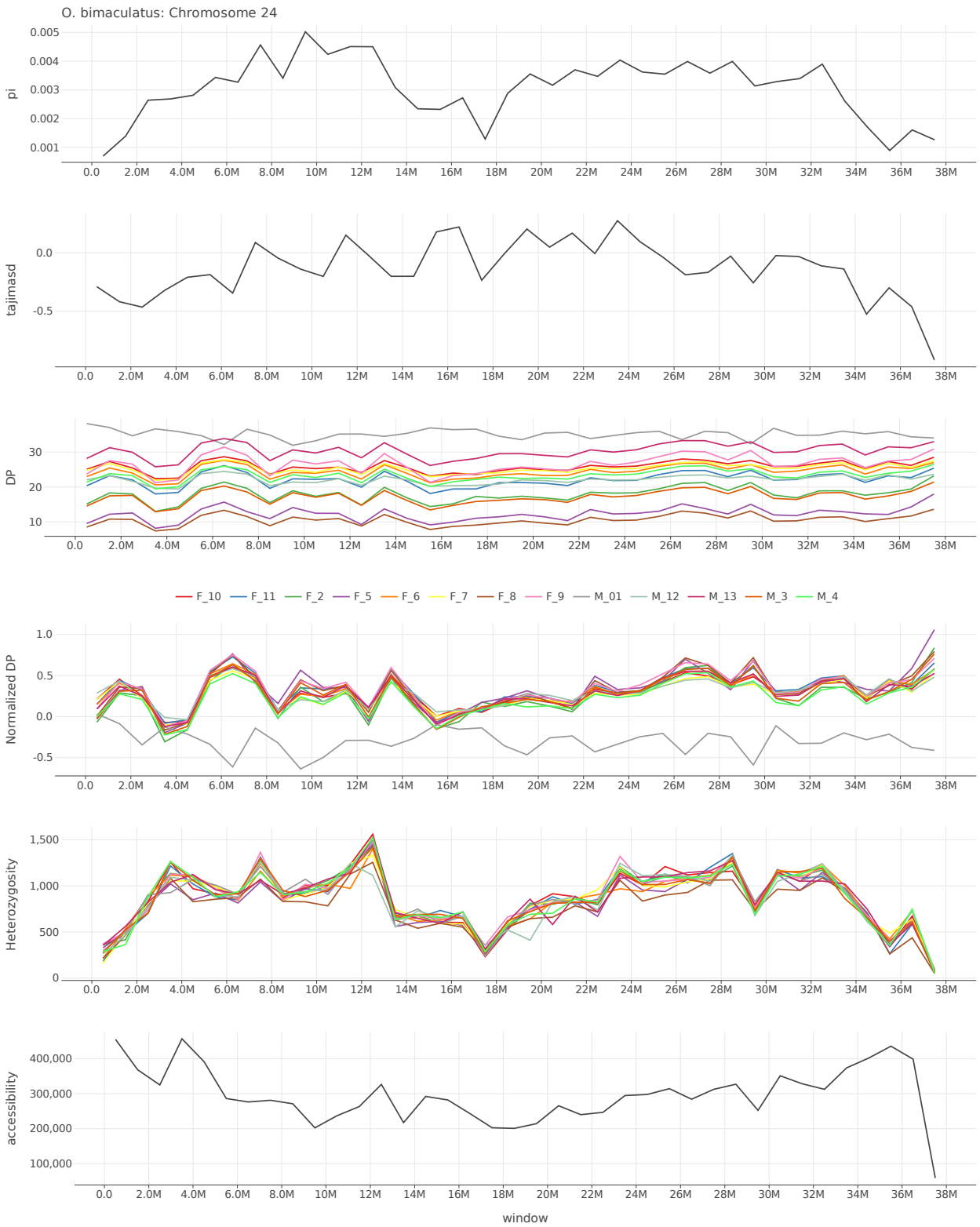

Figure S4.56: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 24 of *O. bimaculatus.*

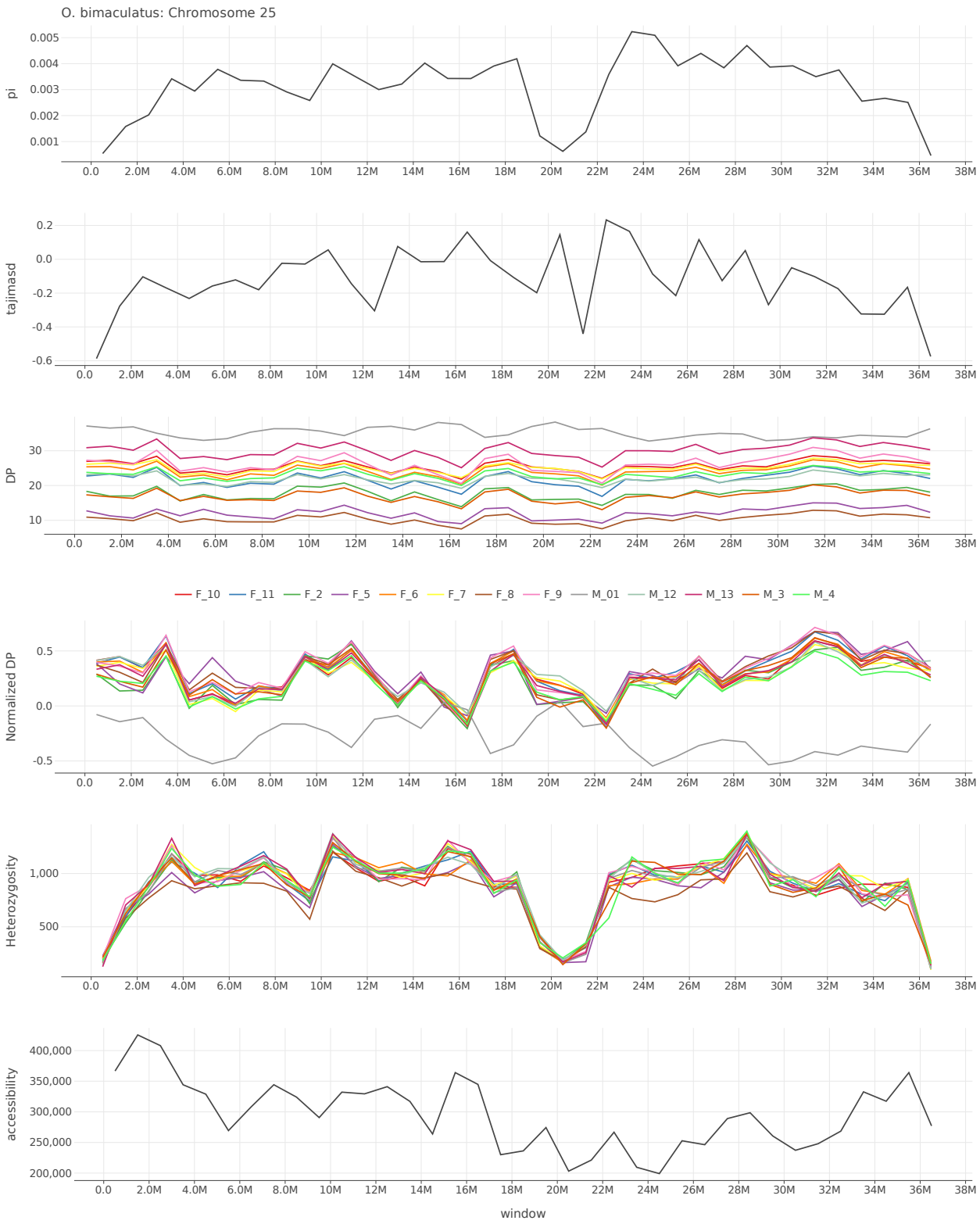

Figure S4.57: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 25 of *O. bimaculatus.*

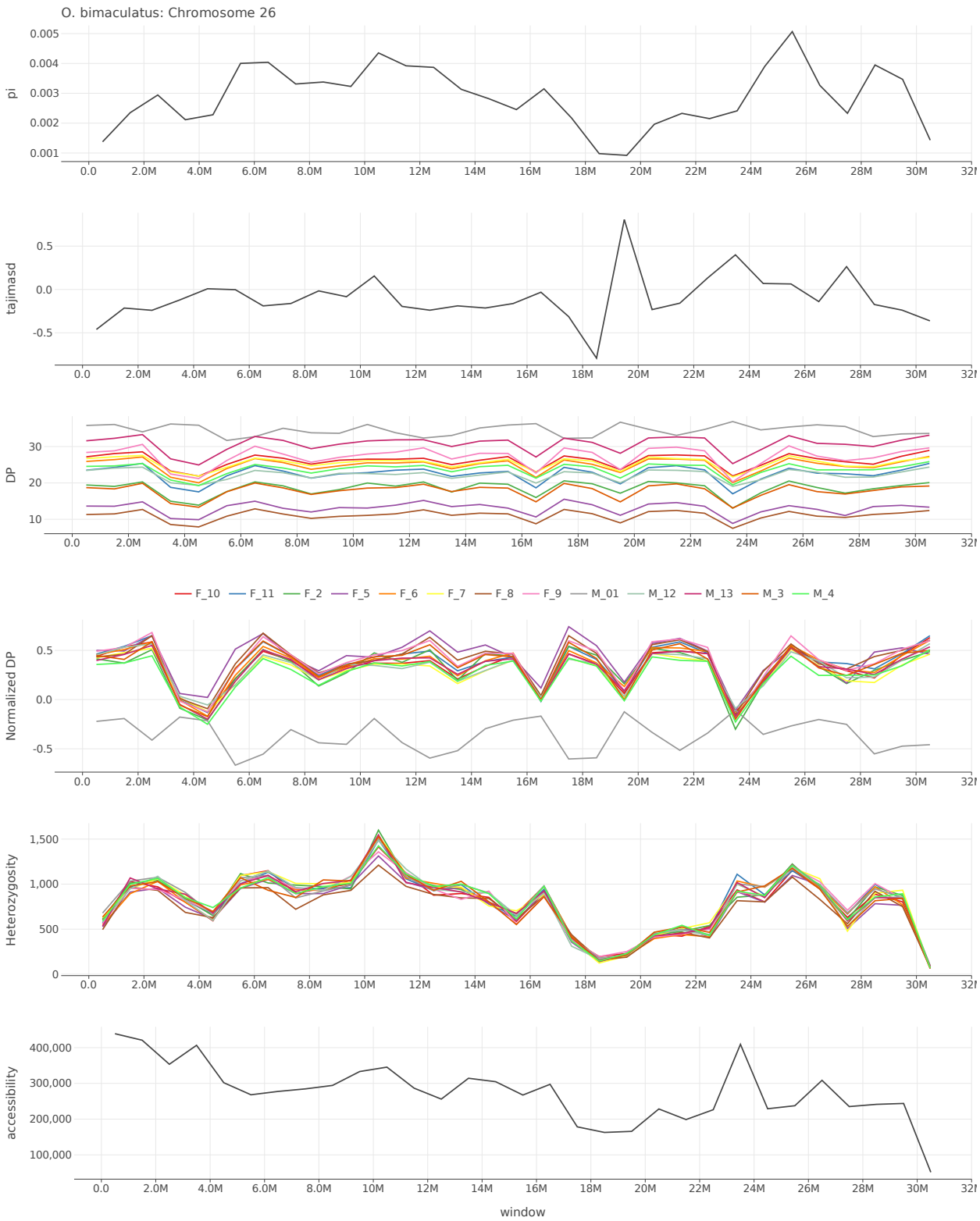

Figure S4.58: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 26 of *O. bimaculatus.*

Figure S4.59: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 27 of *O. bimaculatus.*

Figure S4.60: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 28 of *O. bimaculatus.*

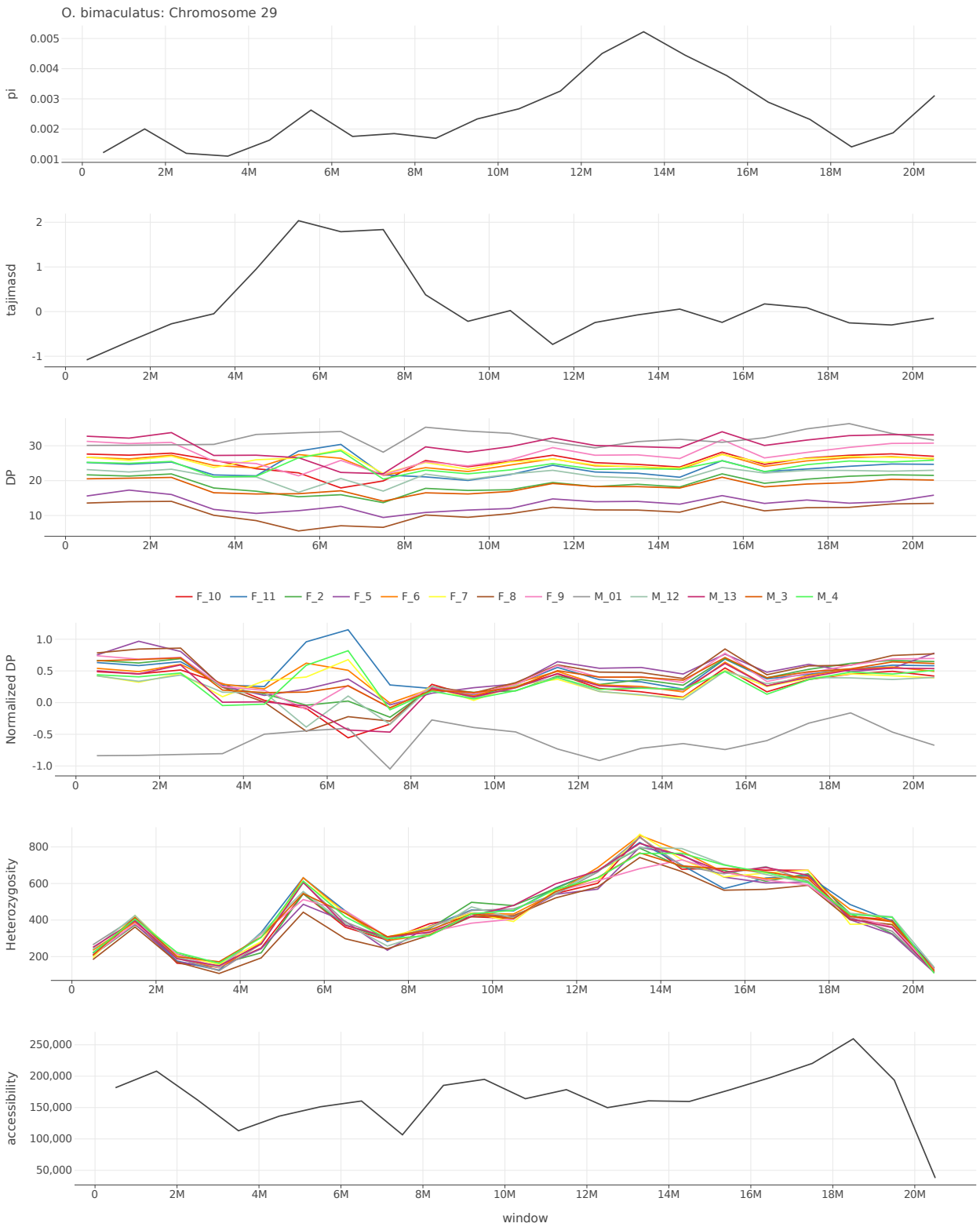

Figure S4.61: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 29 of *O. bimaculatus.*

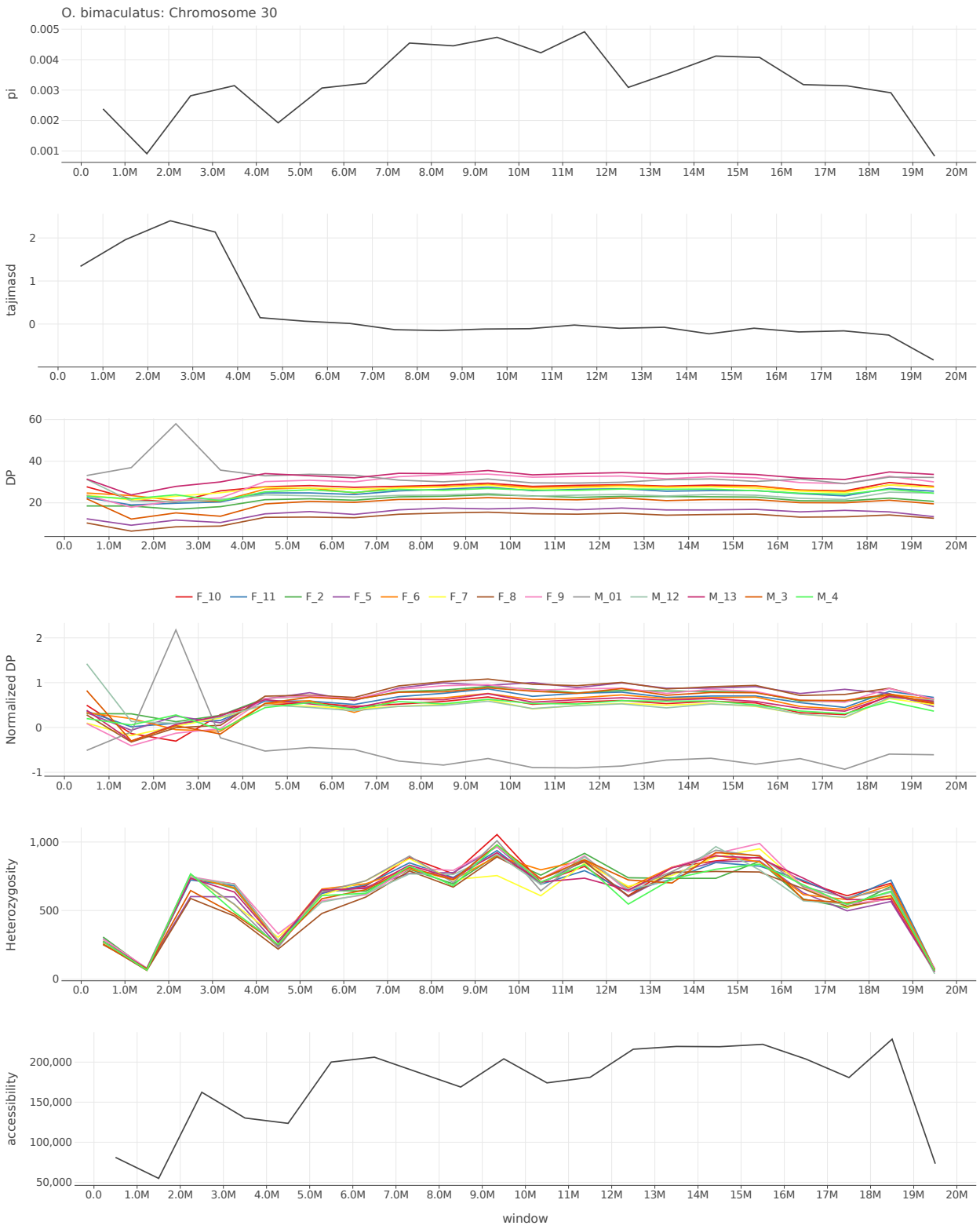

Figure S4.62: Pi, Tajima's D, depth, normalized depth, heterozygosity, and accessibility across chromosome 30 of *O. bimaculatus.*

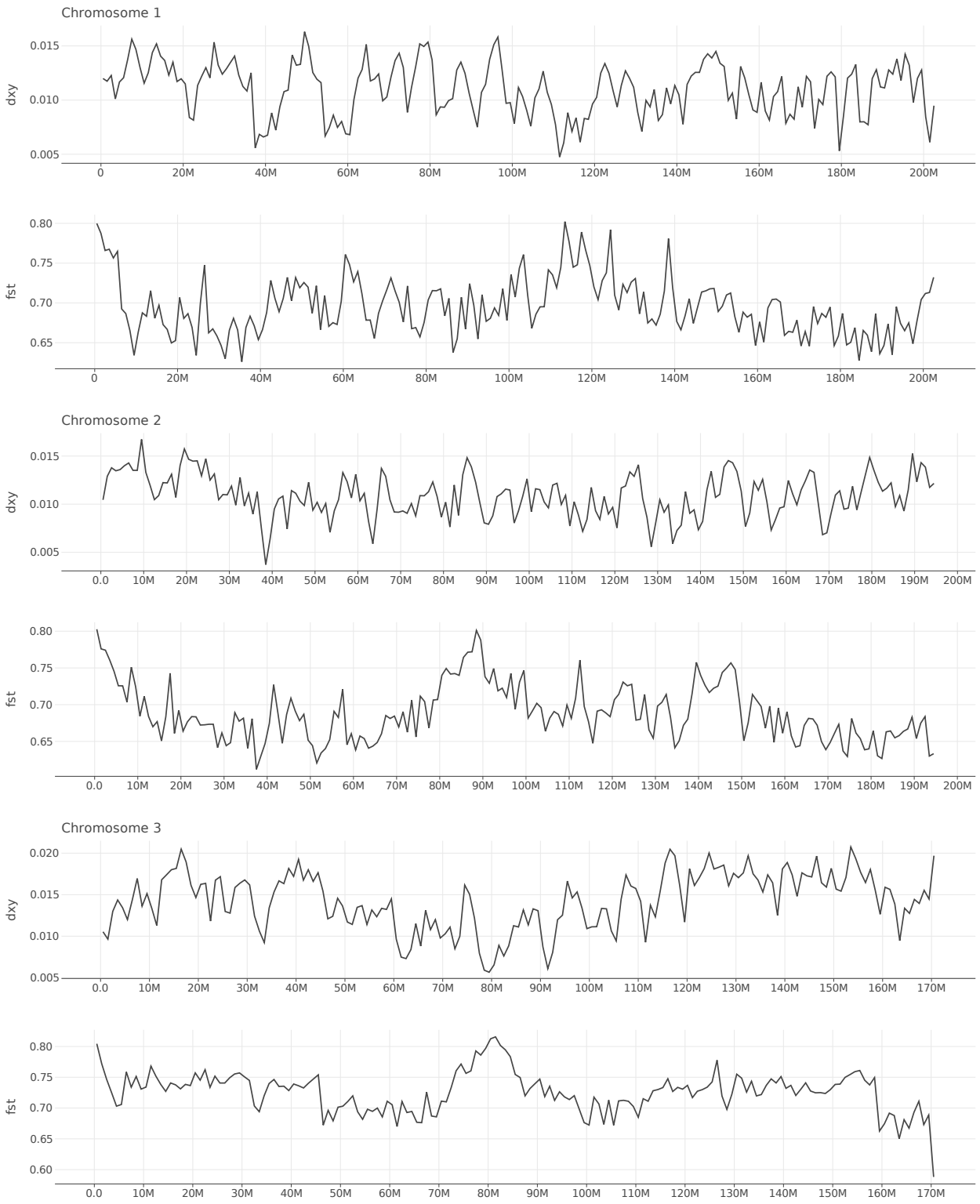

Figure S4.63: d_{XY} and F_{ST} between *O. bimaculoides* and *O. bimaculatus* along chromosomes 1-3.

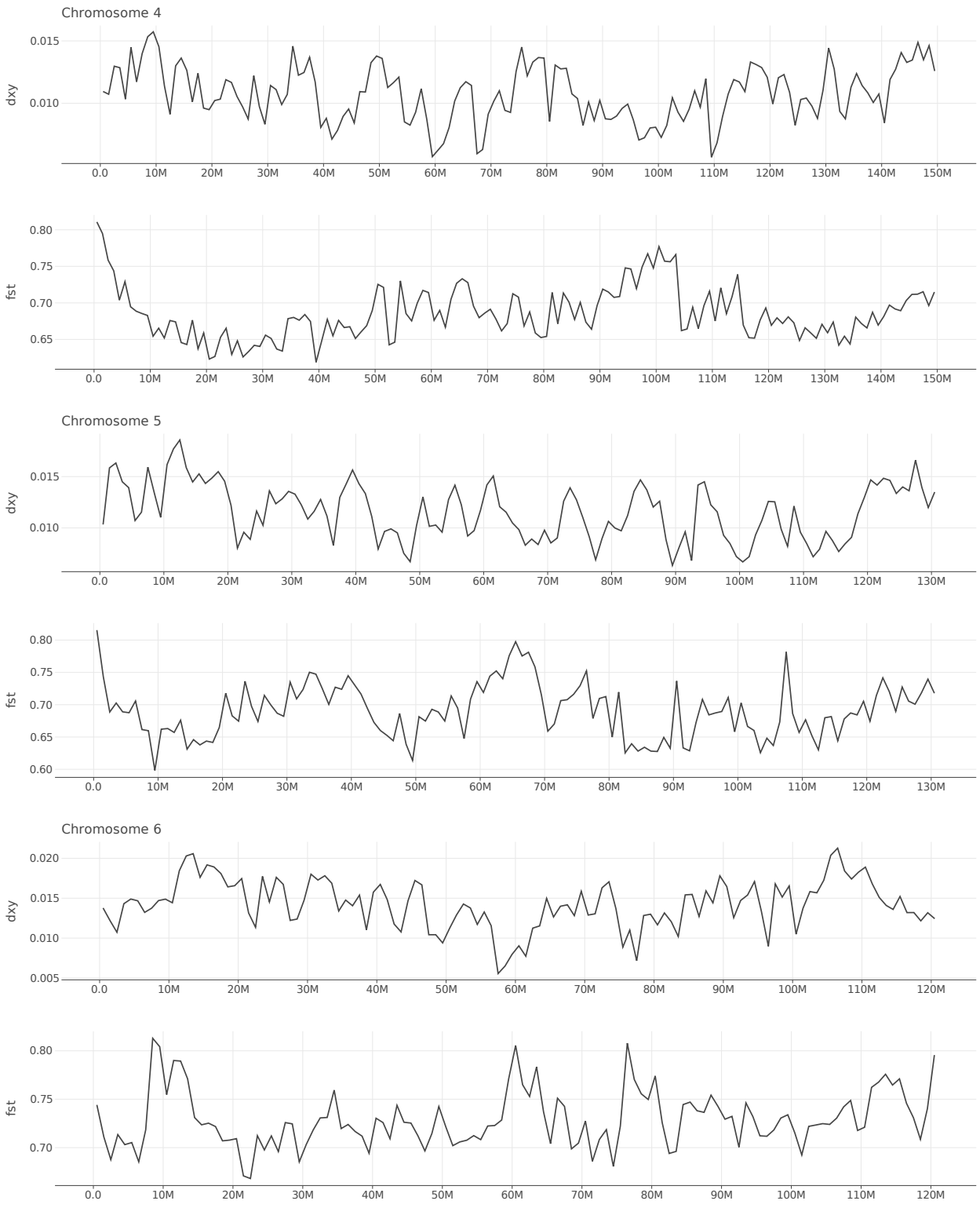

Figure S4.64: d_{XY} and F_{ST} between *O. bimaculoides* and *O. bimaculatus* along chromosomes 4-6.

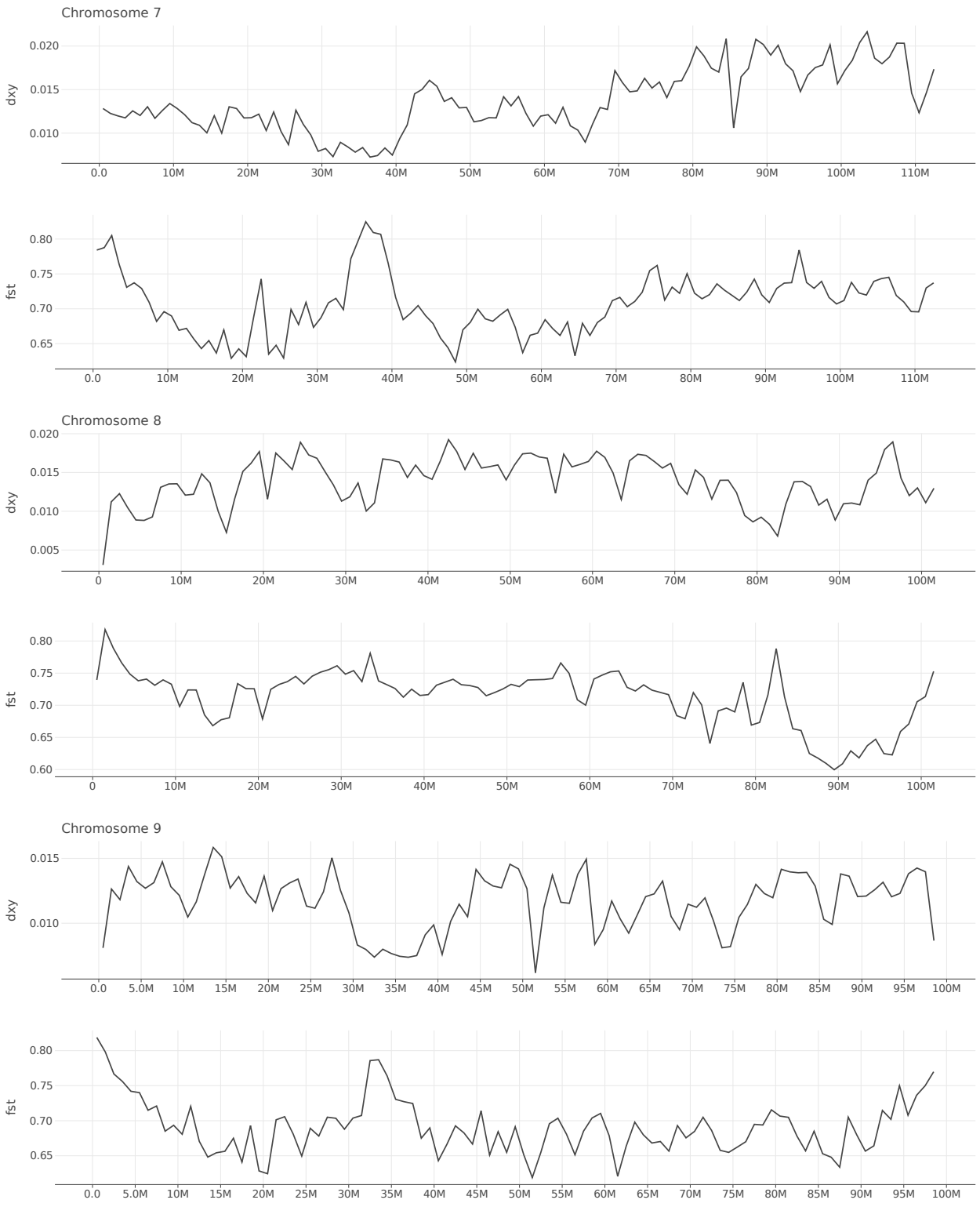

Figure S4.65: d_{XY} and F_{ST} between *O. bimaculoides* and *O. bimaculatus* along chromosomes 7-9.

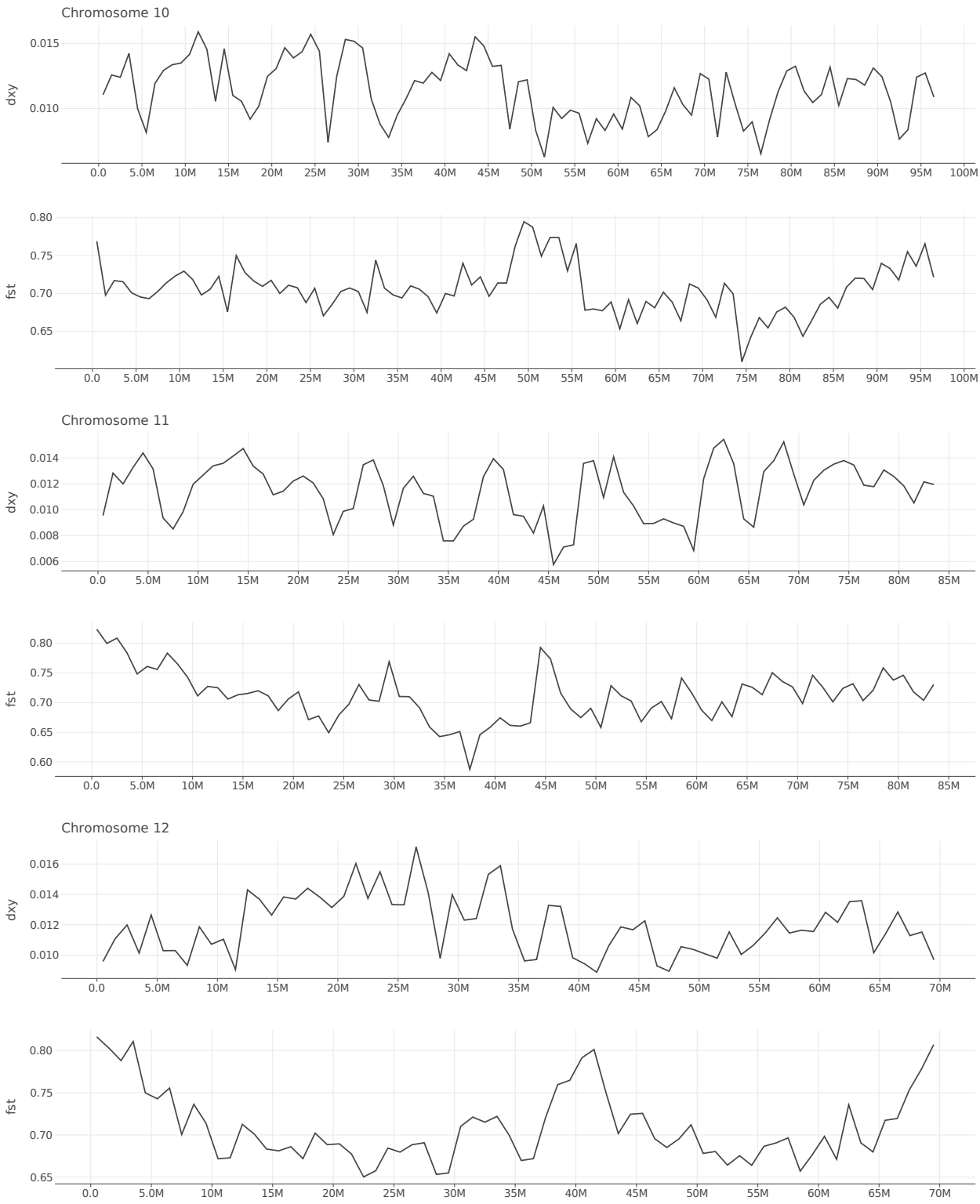

Figure S4.66: d_{XY} and F_{ST} between *O. bimaculoides* and *O. bimaculatus* along chromosomes 10-12.

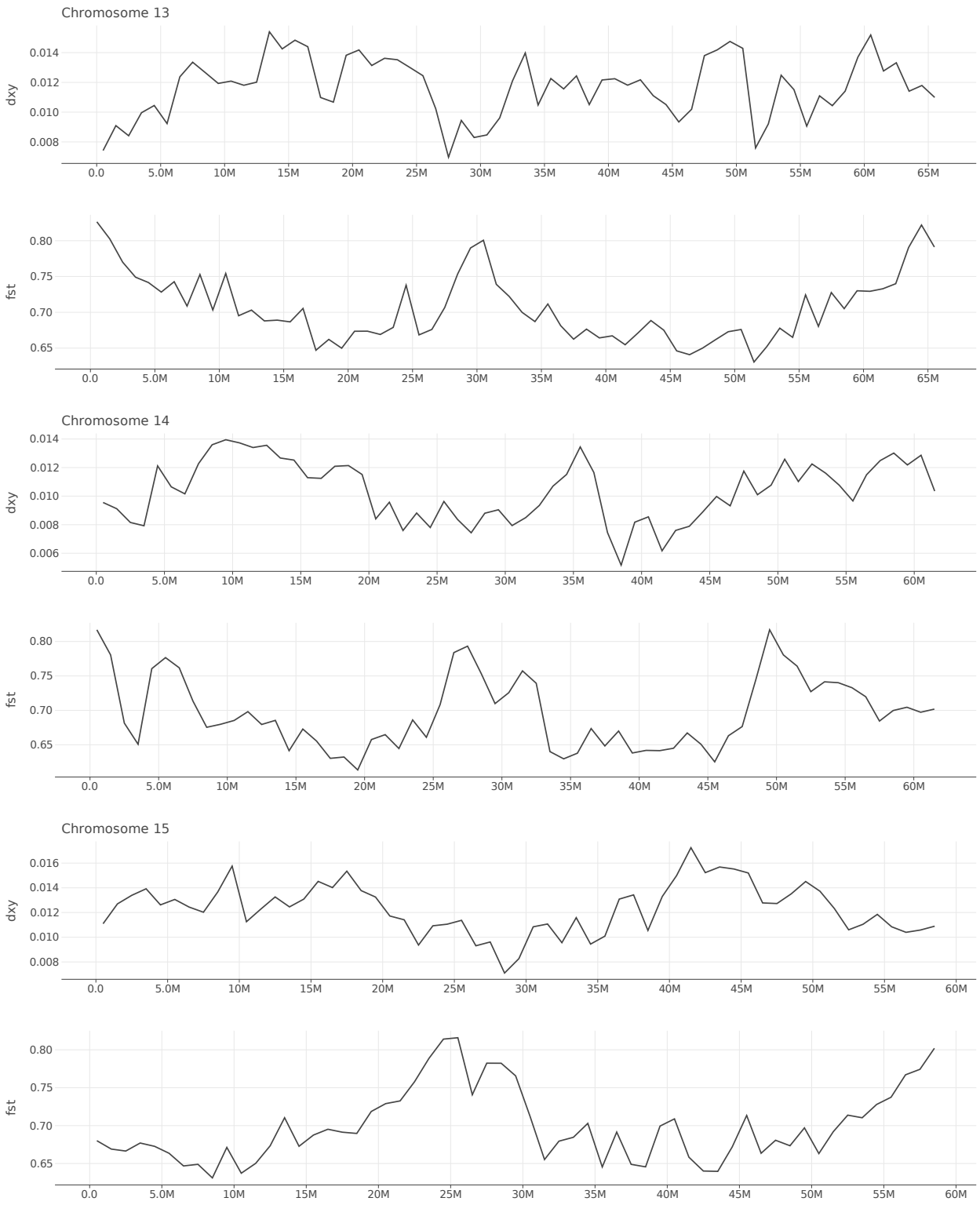

Figure S4.67: d_{XY} and F_{ST} between *O. bimaculoides* and *O. bimaculatus* along chromosomes 13-15.

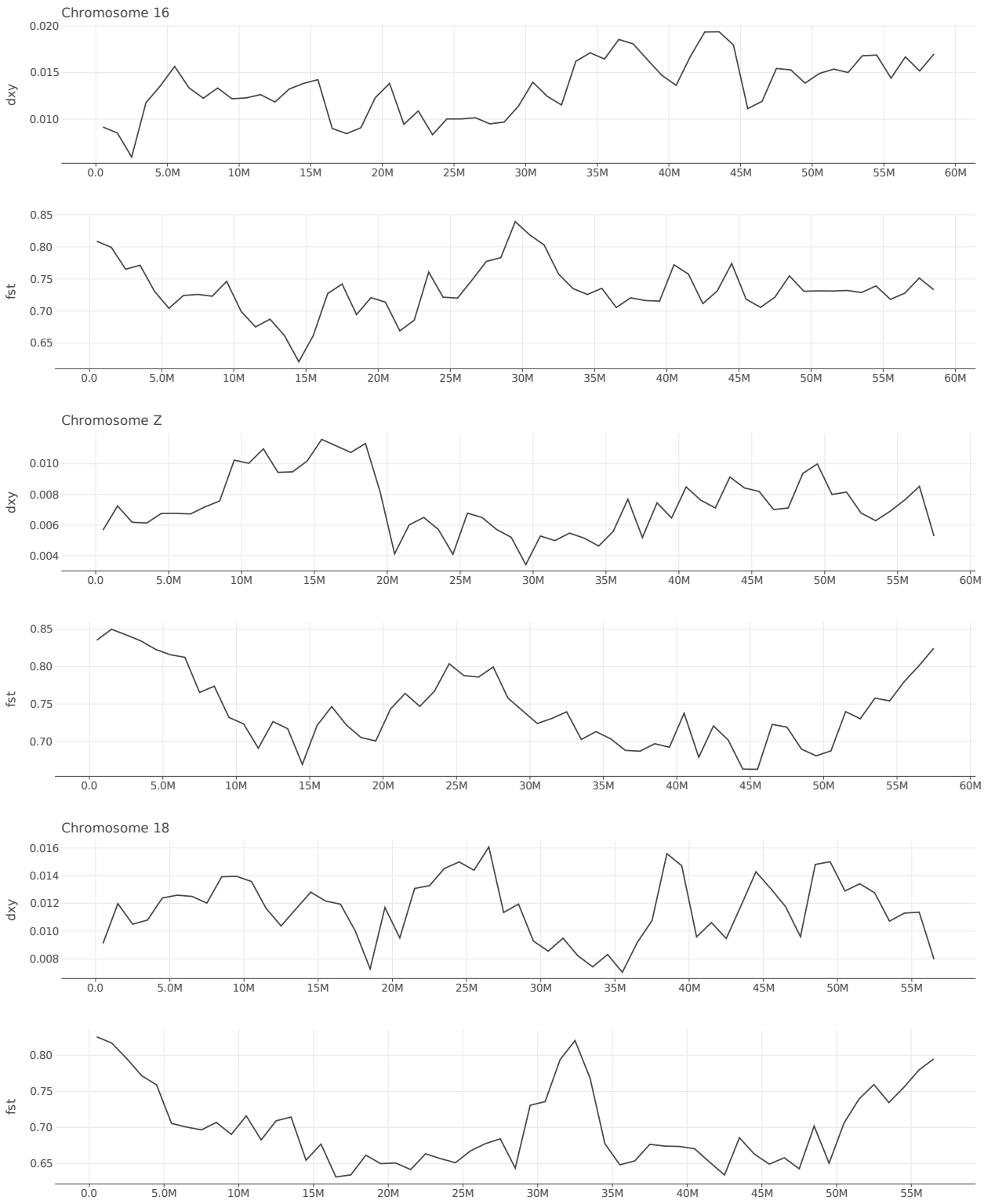

Figure S4.68: d_{XY} and F_{ST} between *O. bimaculoides* and *O. bimaculatus* along chromosomes 16-18.

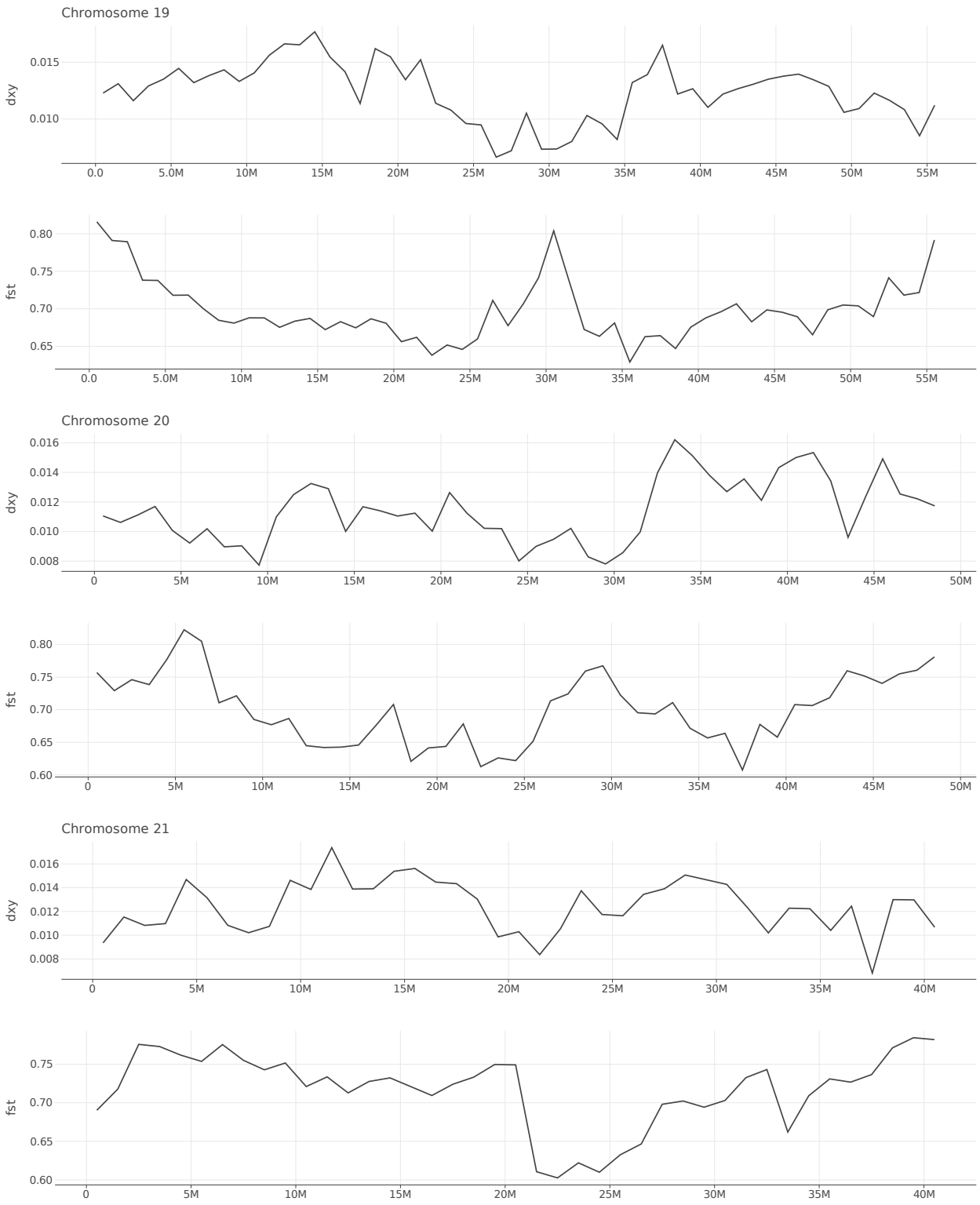

Figure S4.69: d_{XY} and F_{ST} between *O. bimaculoides* and *O. bimaculatus* along chromosomes 19-21.

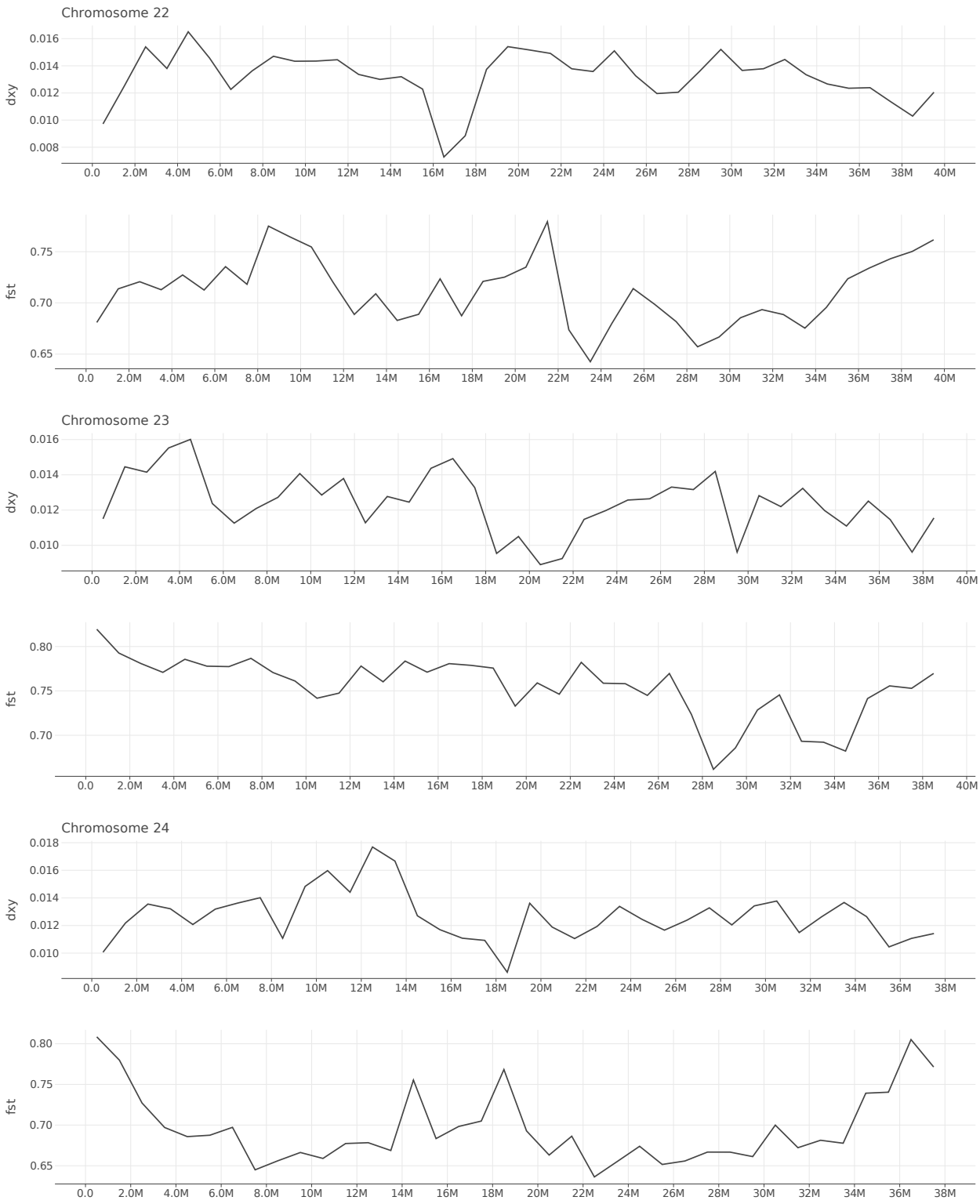

Figure S4.70: d_{XY} and F_{ST} between *O. bimaculoides* and *O. bimaculatus* along chromosomes 22-24.

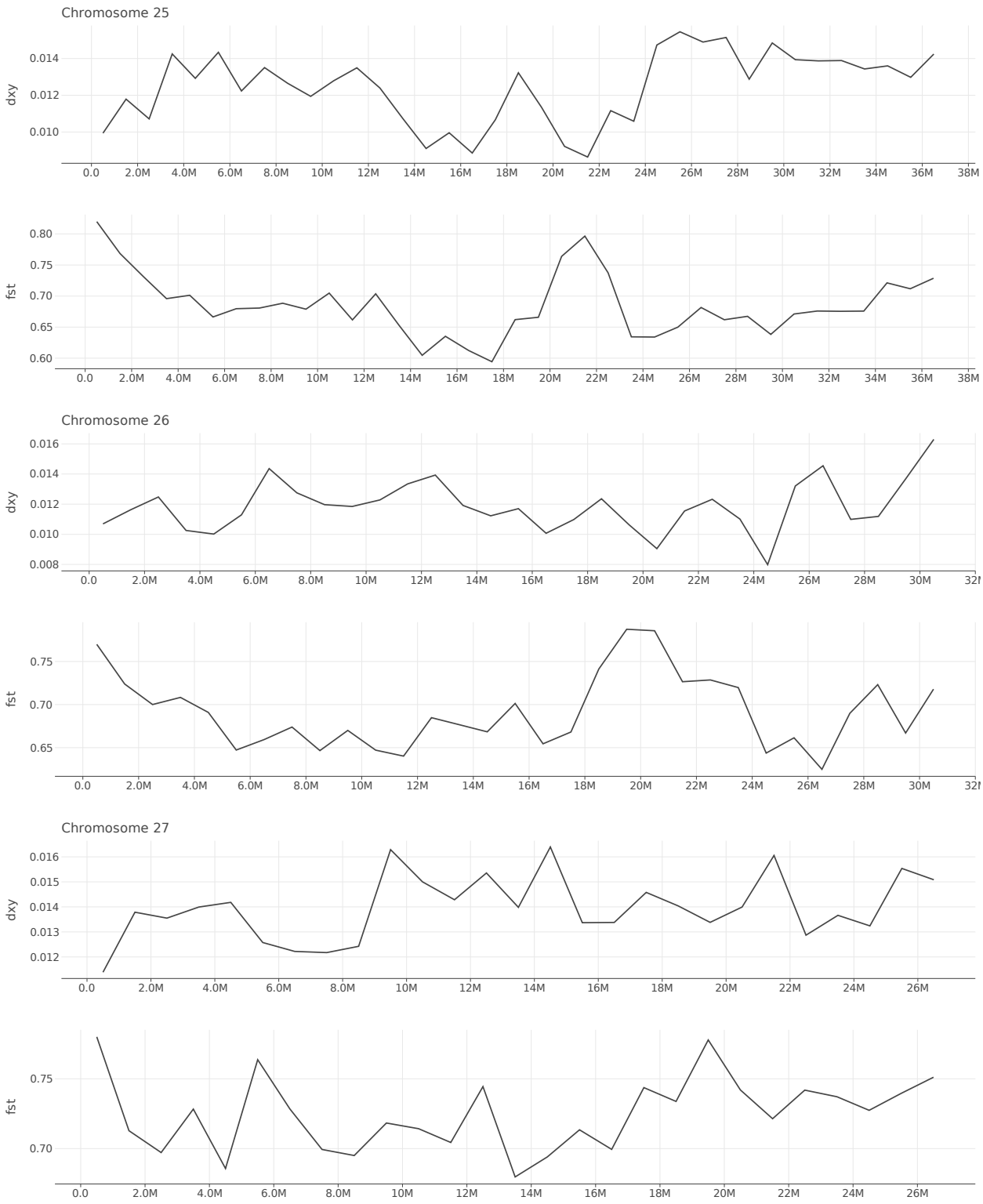

Figure S4.71: d_{XY} and F_{ST} between *O. bimaculoides* and *O. bimaculatus* along chromosomes 25-27.

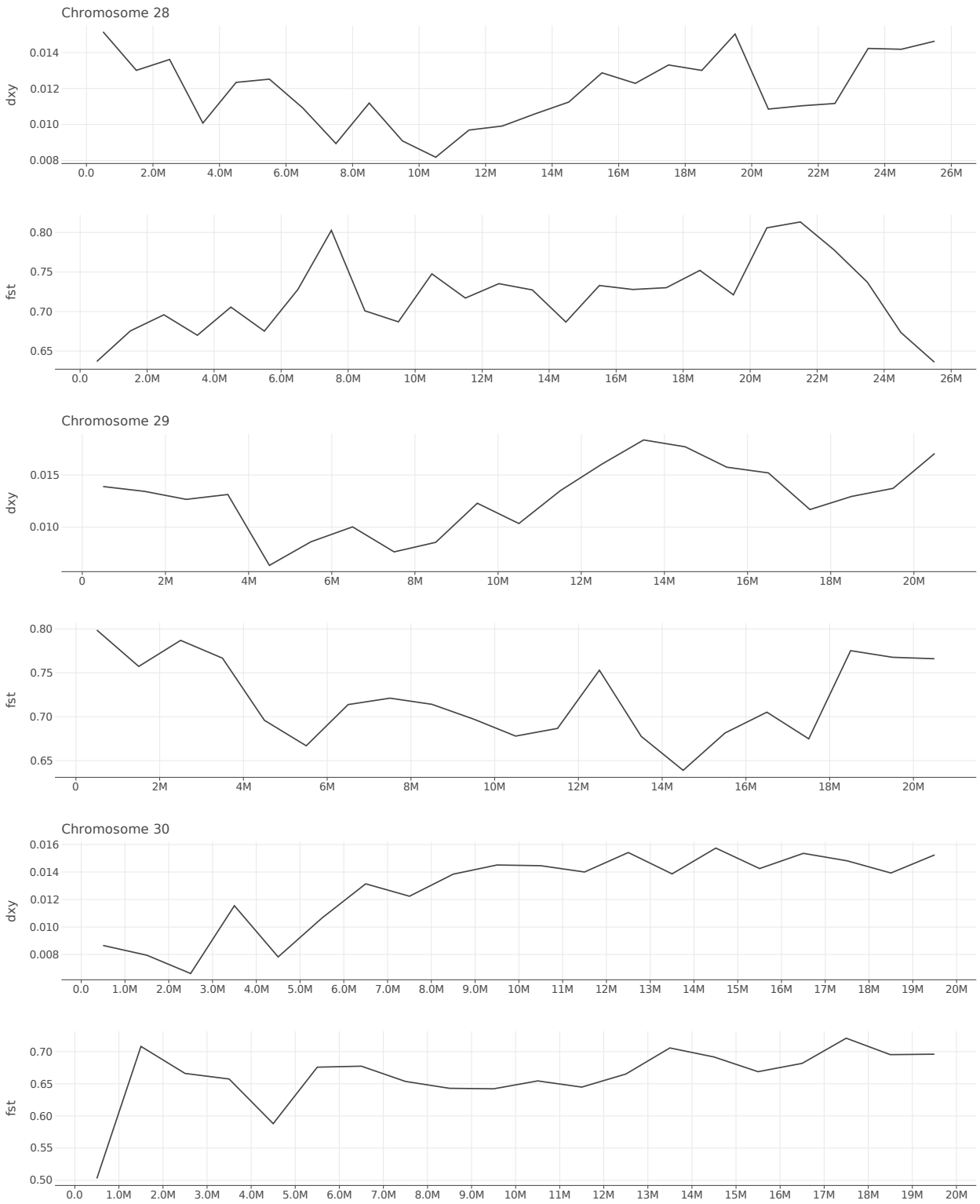

Figure S4.72: d_{XY} and F_{ST} between *O. bimaculoides* and *O. bimaculatus* along chromosomes 28-30.