

Supplemental Information for:

Genomic variation of an endosymbiotic dinoflagellate (*Symbiodinium 'fitti'*) among closely related coral hosts

Hannah G. Reich^{1,2*}, Sheila A. Kitchen^{1,3*}, Kathryn H. Stankiewicz¹, Meghann Devlin-Durante¹, Nicole D. Fogarty⁴, Iliana B. Baums^{1*}

¹Department of Biology, The Pennsylvania State University, University Park, PA 16801 USA

²Present address: Department of Biological Sciences, University of Rhode Island, Kingston, RI 02881, USA

³Present address: Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125

⁴Department of Biology and Marine Biology, Center for Marine Science, University of North Carolina Wilmington, Wilmington, NC 28409

Correspondence: hgreigh16@gmail.com; sak3097@caltech.edu; baums@psu.edu

Table of Contents:

Figure S1	Page 2
Figure S2	Page 3
Figure S3	Page 4
Figure S4	Page 5
Figure S5	Page 6
Figure S6	Page 7
Captions for SI Tables S1-10	Page 8
SI References	Page 9

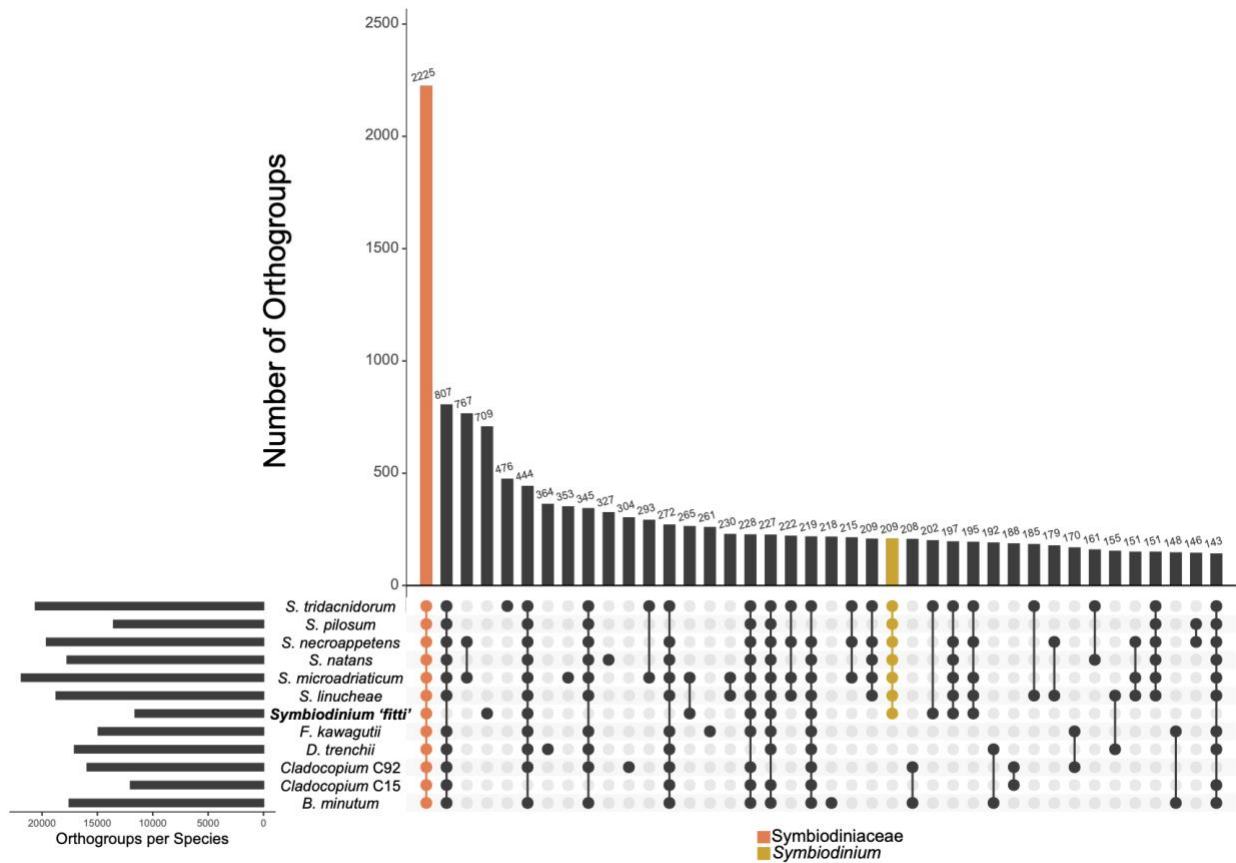


Fig. S1: Upset plot of combinations of gene families shared between *Symbiodinium 'fitti'* and 11 other Symbiodiniaceae assemblies. Gene families shared between *S. 'fitti'* and the 11 other Symbiodiniaceae assemblies are highlighted in coral. The gene families shared between *S. 'fitti'* and the six other assemblies from the genus *Symbiodinium* are highlighted in yellow.

MOLECULAR ECOLOGY

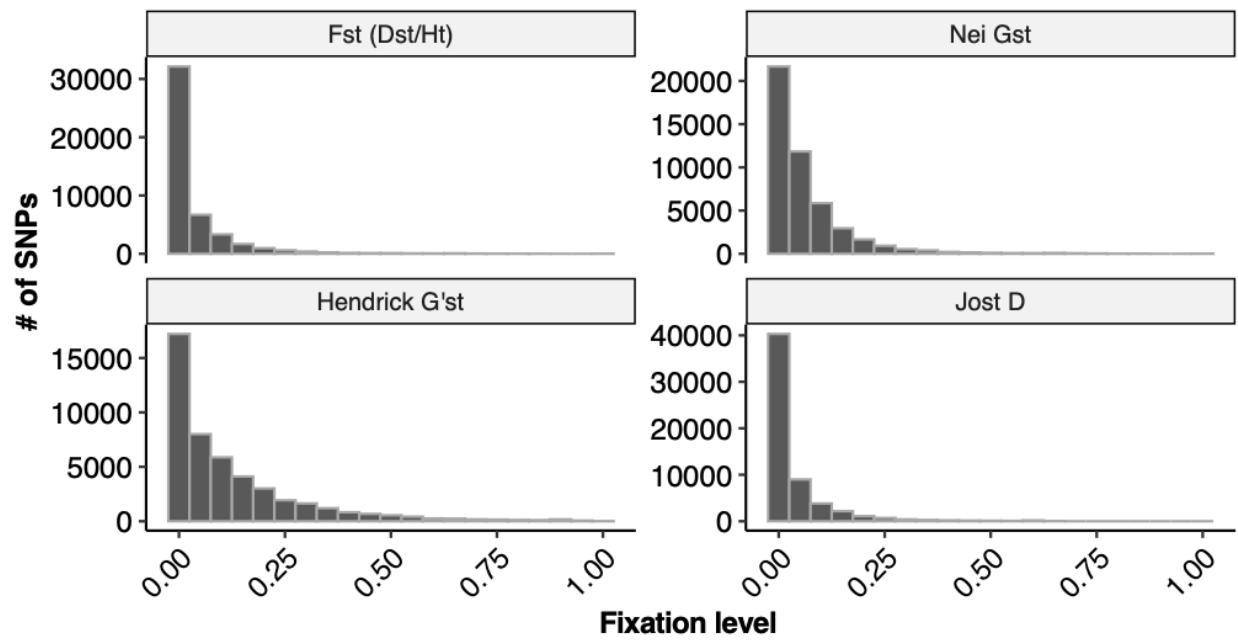


Fig. S2: Histogram of per SNP fixation levels for 58,538 "high quality" SNPs.

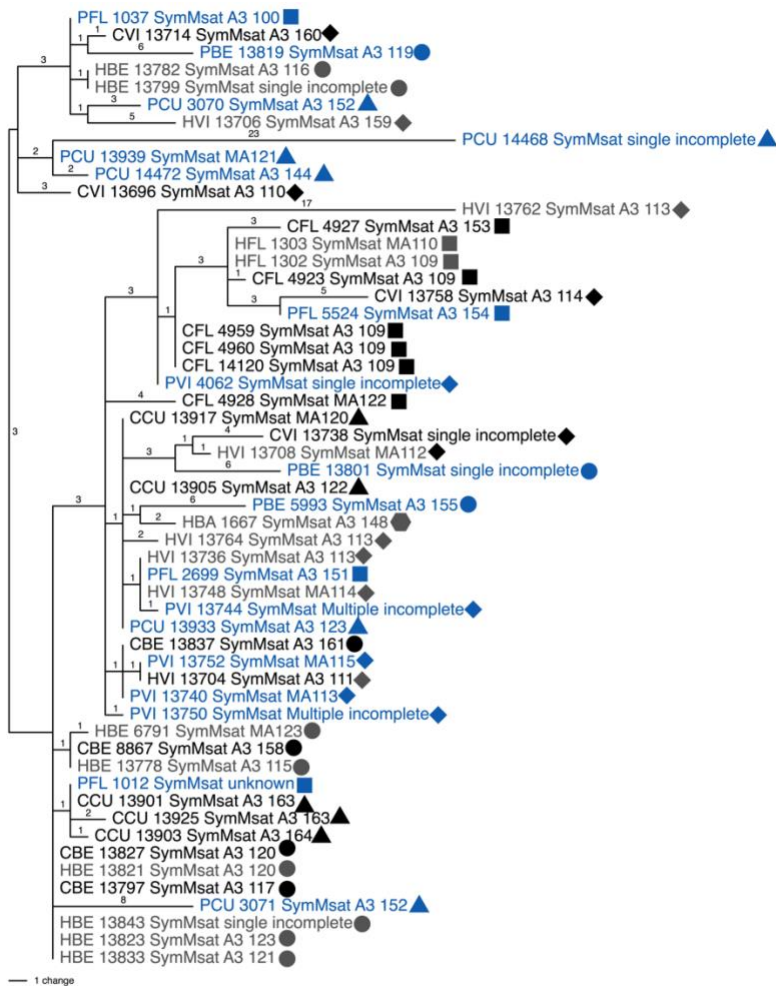


Fig. S3: The lack of differentiation of the *S. 'fitti'* *psbA* non-coding region by host taxa indicates that the genomic (SNP) variation is at the sub-species level. The *psbA* non-coding region is a commonly used marker used to help delimit *Symbiodinium* species, including distinguishing *S. tridacnidorum* (A3, close relative of *S. 'fitti'*) from other species in the genus (Lee et al 2015). Samples from different host taxon are denoted by point color and point shape reflects geographic location. Values refer to the number of base differences that differentiate samples.

5,067 *Symbiodinium 'fitti'* selection outlier SNPs

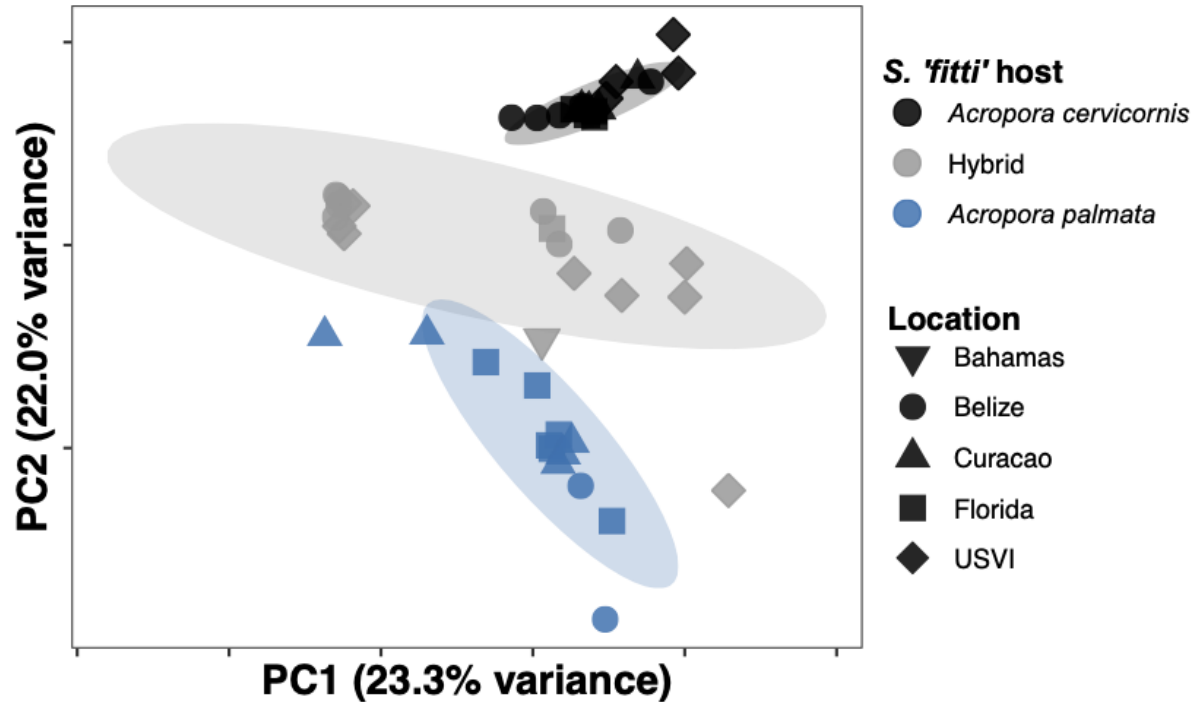


Fig. S4: Principal component analysis (PCA) of 5,067 *Symbiodinium 'fitti'* SNPs under selection illustrates genomic differentiation by host taxon. Ellipses represent 95% confidence intervals for the multivariate distribution of each host taxa. Samples from different host taxon are denoted by point color and point shape reflects geographic location.

16,536 *Symbiodinium 'fitti'* SNPs in gene regions

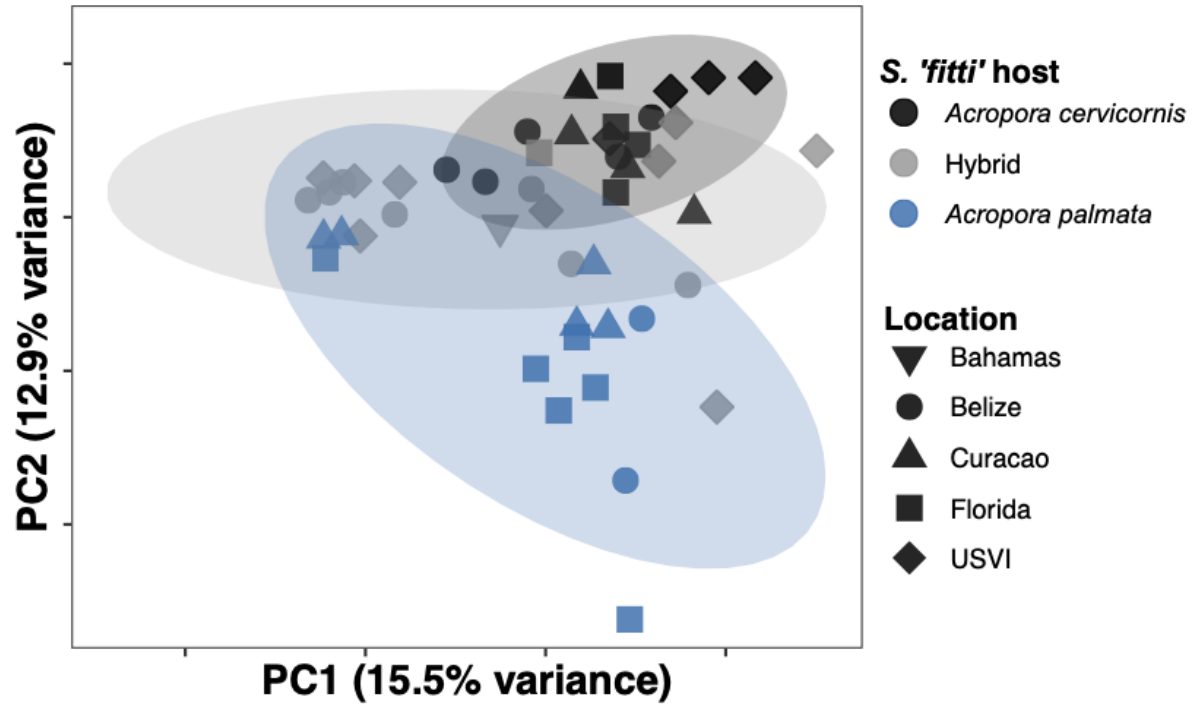


Fig. S5: Principal component analysis (PCA) of 16,536 *Symbiodinium 'fitti'* SNPs in gene regions does not illustrate genomic differentiation by host taxon. Ellipses represent 95% confidence intervals for the multivariate distribution of each host taxa. Samples from different host taxon are denoted by point color and point shape reflects geographic location.

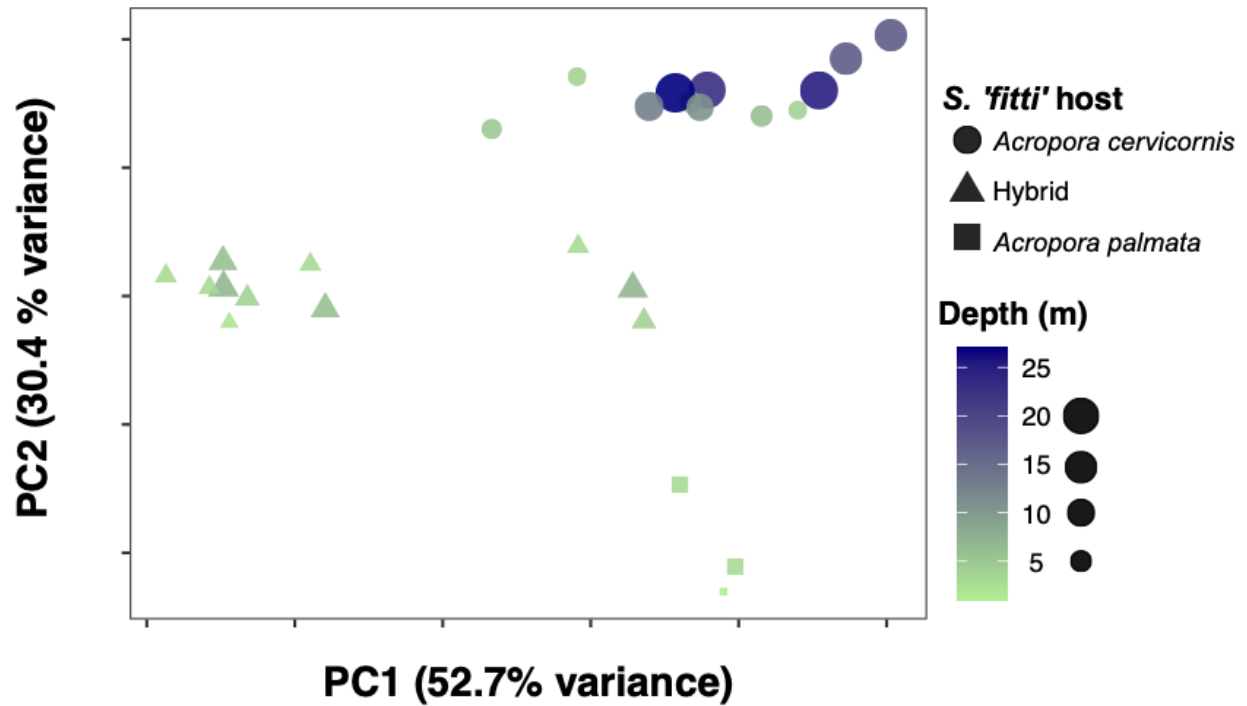


Fig. S6: Principal component analysis (PCA) of 58,583 genotyping *Symbiodinium 'fitti'* SNPs from 26 samples with depth information. Samples from different host taxon are denoted by point shape whereas point size and color reflect sampling depth.

Captions for SI Tables S1-10*

Table S1: Sample information including host taxa, location, sampling depth, sequencing platform, and accession number for all *Symbiodinium 'fitti'* samples. This information includes read counts, mapping rates, and SNP summary statistics. Environmental information is provided for a subset of samples (sampling depth, benthic substrate, average *Acropora palmata*, *A. cervicornis*, and hybrid abundance). Multi-locus genotypes from 12 microsatellite loci, 58,538 “high quality” SNPs, and 6,813 “conservative” SNPs with no missing data are included.

Table S2: Genome assembly summary statistics and gene metrics for *Symbiodinium 'fitti'* sample PFL14120, *Symbiodinium microadriaticum* (Aranda et al 2016) and *S. tridacnidorum* (Shoguchi et al 2018). *S. microadriaticum* and *S. tridacnidorum* gene metrics are from Chen et al 2020.

Table S3: Annotation information for the 37,000+ genes in *Symbiodinium 'fitti.'* Predicted genes were queried against ncbi, uniprot, trembl, and Symbiodiniaceae databases

Table S4: Annotation information and selection outlier (BayeScan, PCAdapt) statistics for 16,536 *Symbiodinium 'fitti'* SNPS that are predicted to fall in coding regions.

Table S5: AMOVA of indicates that 11.6% of *Symbiodinium 'fitti'* genomic variation is due to host taxon whereas no variation (-5.6%) is due to location.

Table S6: Summary statistics from Structure Harvester including the Evanno's *delta K* method which predicted three main clusters.

Table S7: Selection outlier statistics and per SNP summary statistics loci identified as selection outliers by BayeScan (370) and PCAdapt (4,987). 307 selection outlier SNPs were shared between the two programs.

Table S8: Summary statistics for the predicted downstream effects of SNPs (generated by the SnpEff program). 61,259 (82.9%) modifier/non-coding, 4,670 (6.3%) moderate/mostly harmless, 7,825 (10.6%) low impact variants that might change protein efficiency/effectiveness, and 128 (0.17%) highly disruptive SNPs were identified. Of these mutations, the majority (61.9%) are predicted as silent mutations (7,718) a large portion (37.6%) are predicted as non-synonymous (missense) mutations (4,692), and a small fraction (0.54%) are predicted as nonsense mutations (68). These mutations are predicted to cause 12,472 codon changes and 4,760 amino acid changes.

Table S9: Predicted downstream effects (SnpEff) and annotation information for 16,563 *Symbiodinium 'fitti'* SNPs in coding regions.

Table S10: Predicted downstream effects (SnpEff) and annotation information for 65 SNP sites in coding regions that were identified as selection outliers by both PCAdapt and BayeScan. Within the 65 SNP sites, SnpEff predicted 4,053 downstream effects among the 47 *S. fitti* samples. The majority (3,813/4,053) of these predicted effects occurred in ~30 positions of a RNA-directed DNA polymerase from mobile element jockey (Trmbl ID = A0A2B4S3A8_STYPI, Trmbl e-value = 3.51E-16; NCBI nr ID = XP_022792420.1, NCBI nr e-value = 2.89E-19) from a *Stylophora pistillata* metagenome (also *Symbiodinium natans* gene ID = Snat_CCMP2548.gene28935.mRNA1, E-value = 0). It is highly unlikely that these SNPs are a source of host contamination because *Symbiodinium fitti* scaffold's (scaffold19374|size5297) GC content is 51% which matches other *Symbiodinium* assemblies but does not match the GC content of the *S. pistillata* conspecific (39%).

*Tables S1-S10 are in File S2 (excel document): mec15952-sup-0002-TableS1-S10.xlsx

SI References

- Aranda M., Li Y., Liew Y. J., Baumgarten S., Simakov O., Wilson M. C., . . . Voolstra C. R. (2016) Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific reports* **6**, 39734.
- Chen Y., González-Pech R. A., Stephens T. G., Bhattacharya D., & Chan C. X. (2020) Evidence That Inconsistent Gene Prediction Can Mislead Analysis of Dinoflagellate Genomes. *Journal of Phycology* **56**, 6-10.
- Shoguchi E., Beedessee G., Tada I., Hisata K., Kawashima T., Takeuchi T., . . . Shinzato C. (2018) Two divergent *Symbiodinium* genomes reveal conservation of a gene cluster for sunscreen biosynthesis and recently lost genes. *BMC genomics* **19**, 458.