

Supplementary Materials for

**Spatially varying selection between habitats drives physiological shifts and local adaptation in a broadcast spawning coral on a remote atoll in Western Australia**

Luke Thomas\*, Jim N. Underwood, Noah H. Rose, Zachary L. Fuller, Zoe T. Richards, Laurence Dugal, Camille M. Grimaldi, Ira R. Cooke, Stephen R. Palumbi, James P. Gilmour

\*Corresponding author. Email: [l.thomas@aims.gov.au](mailto:l.thomas@aims.gov.au)

Published 27 April 2022, *Sci. Adv.* **8**, eabl9185 (2022)  
DOI: [10.1126/sciadv.abl9185](https://doi.org/10.1126/sciadv.abl9185)

**The PDF file includes:**

Figs. S1 to S12  
Tables S10 and S11  
Legends for tables S1 to S9

**Other Supplementary Material for this manuscript includes the following:**

Tables S1 to S9

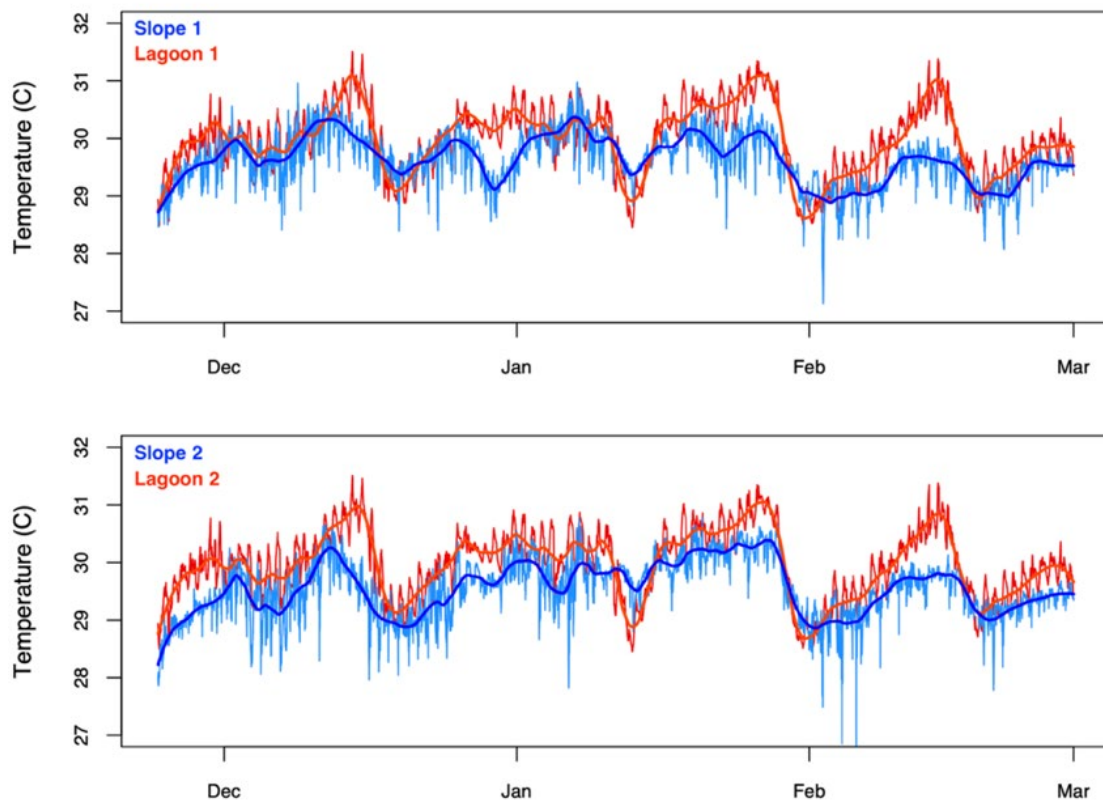


fig. S1 Time-series temperature data (10-minute intervals) from lagoon and slope habitats at Clerke Reef collected across the 2017/2018 summer month. See ESM2 for fine-scale resolution across a two-week window in 2019.

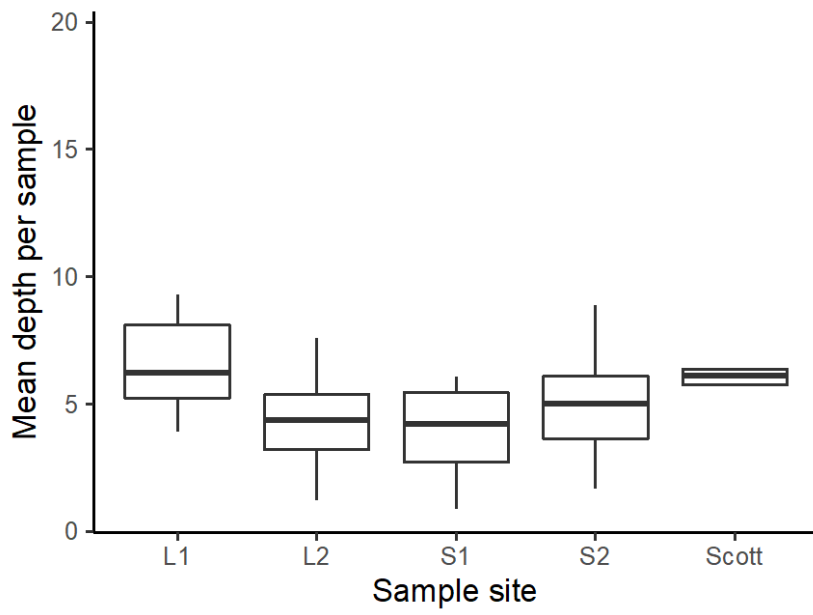


fig. S2 Boxplot of mean coverage per sample for each sample site.

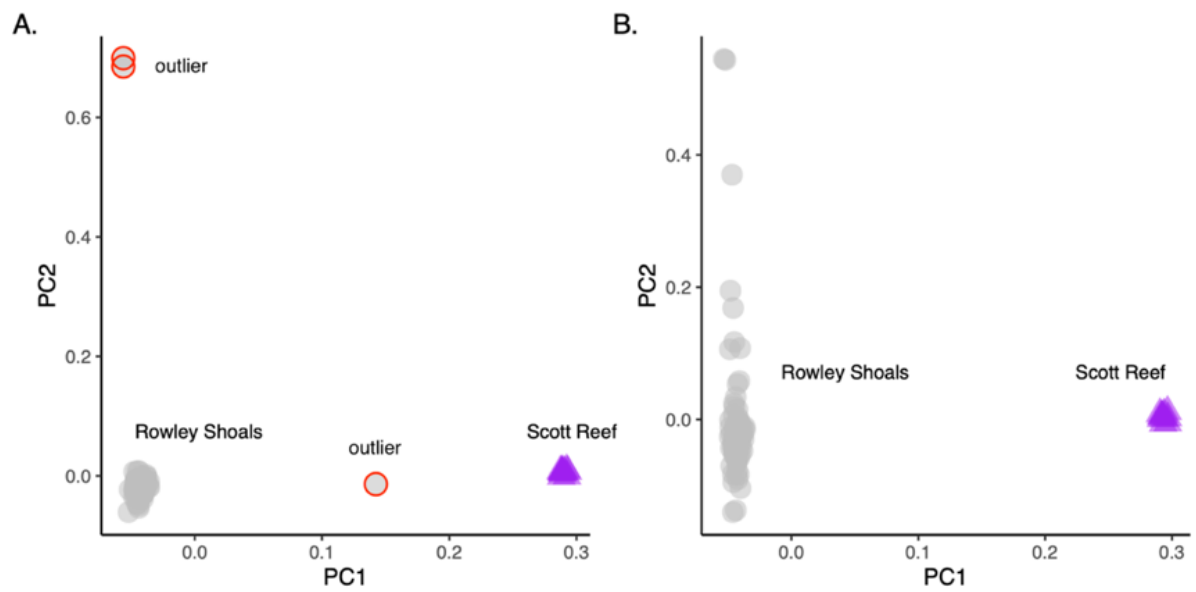


fig. S3 Initial screening of individuals using Principal Component Analysis in PCangsd. Plots display the first two principal components component in all samples (A) and across samples once outliers (red outline) were removed from the dataset (B). Samples cluster strongly by reef system (Rowley Shoals and Scott Reef).

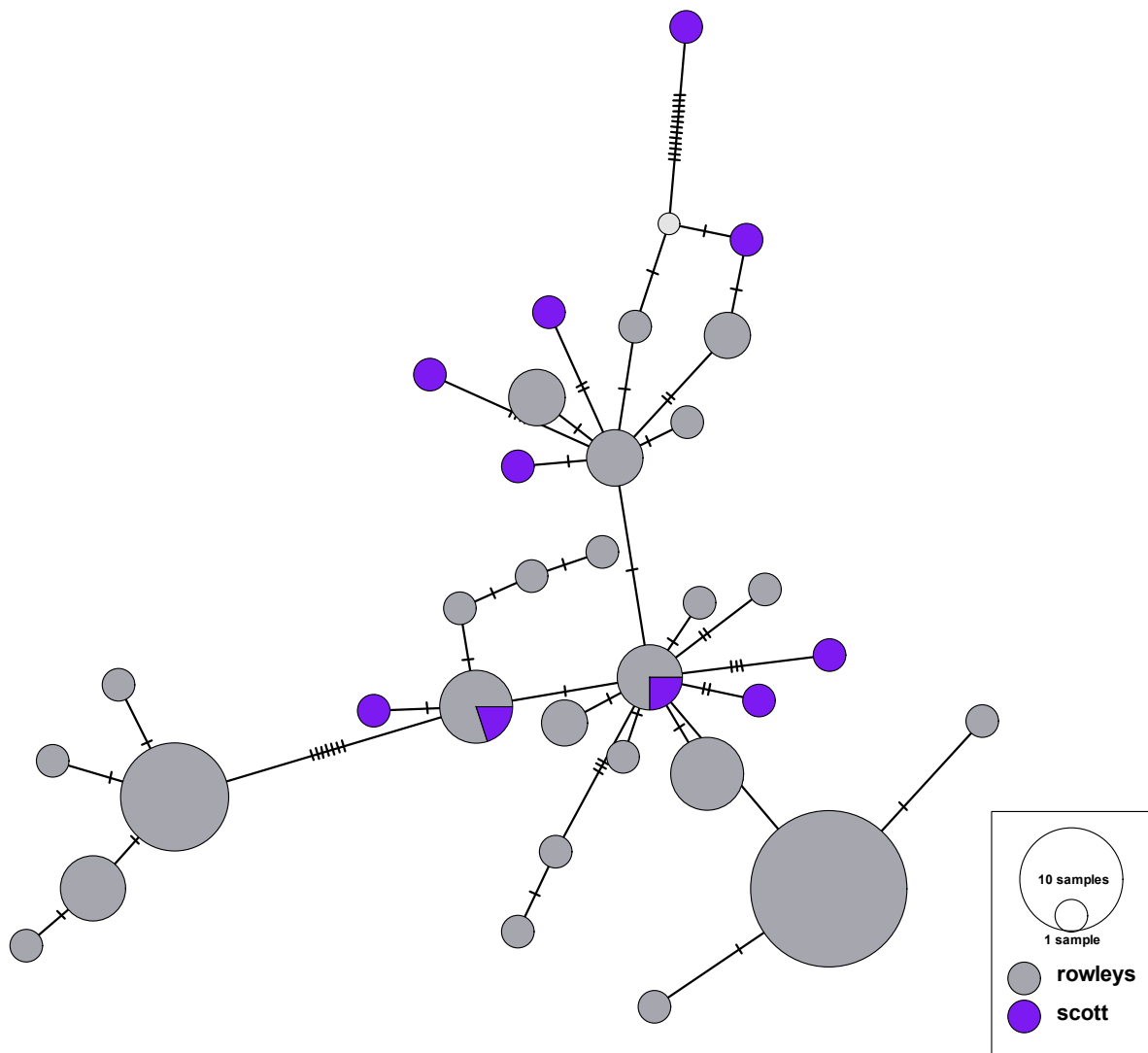


fig. S4 Sequence reads were also mapped to the complete *A. tenuis* mitochondrion and achieved a mean coverage of 313x (+\ 18.27 SE) per sample. TCS network based on extracted consensus sequences for each sample across the complete mitochondrion

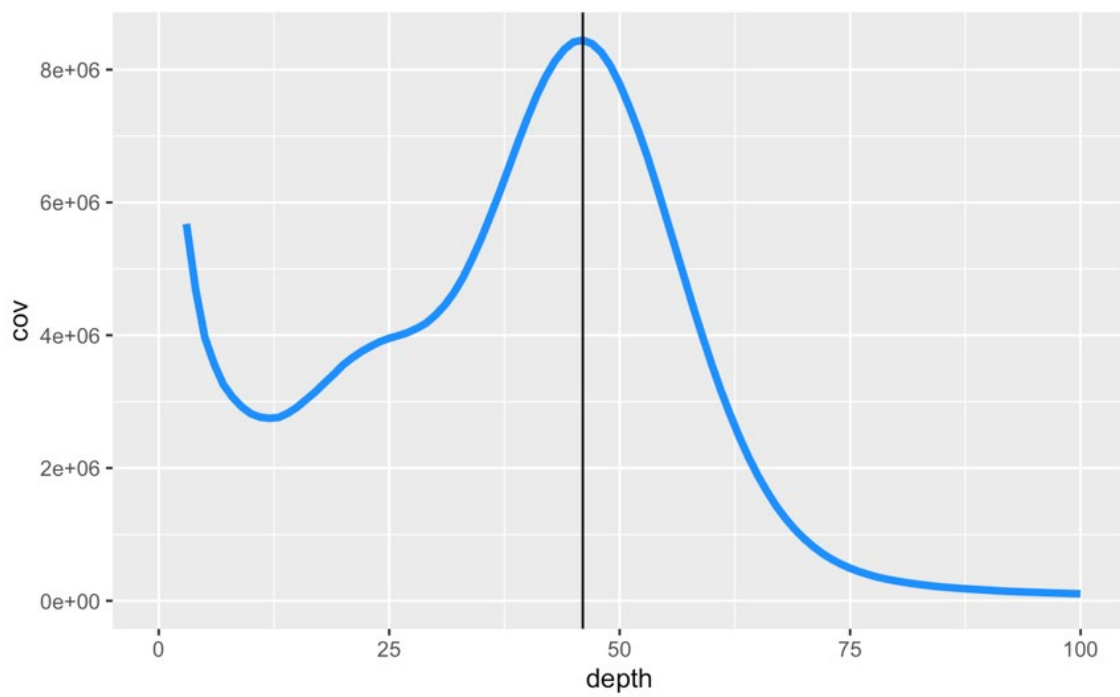


fig. S5 Density plot of sequence coverage per base pair for a single individual from the Rowley Shoals used for MSMC analyses.

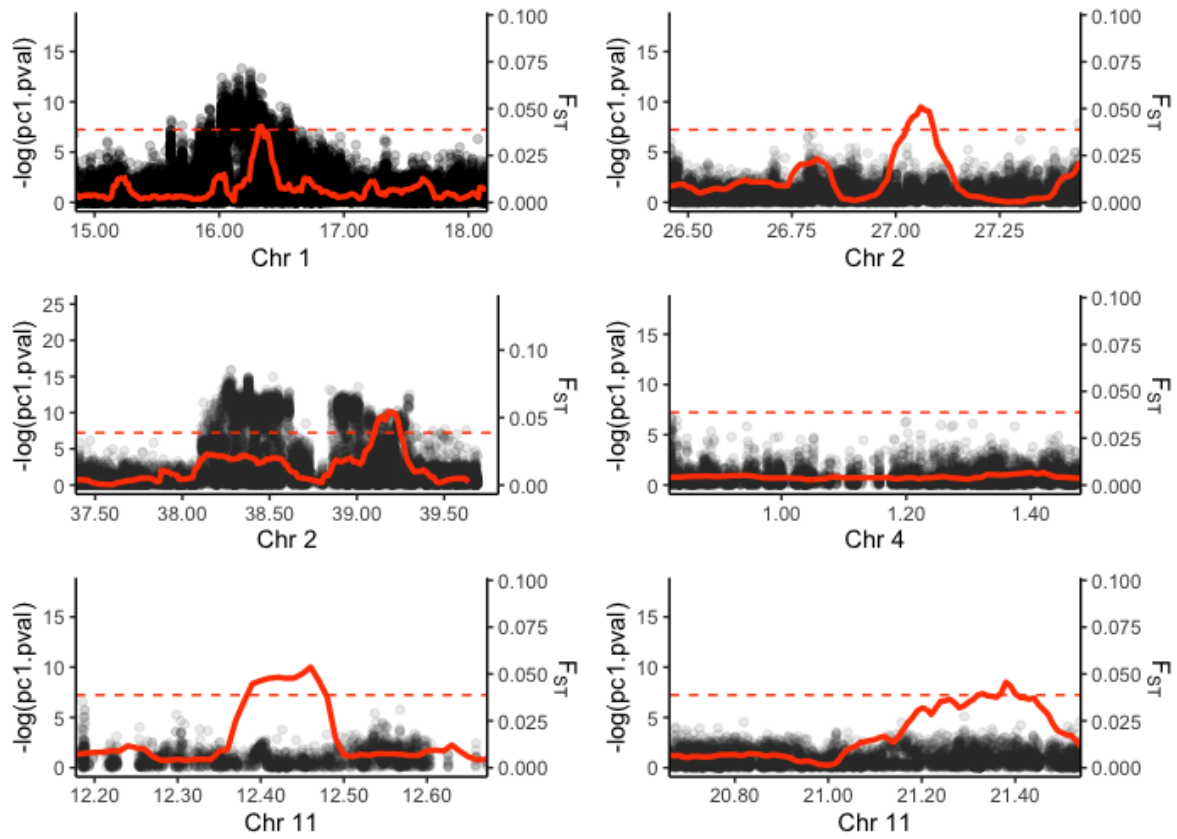


fig. S6 Zoomed in plots of  $F_{ST}$  (red line-100kb windows) and the population-independent selection coefficients for each SNP (black points) surrounding the 6  $F_{ST}$  outlier regions.

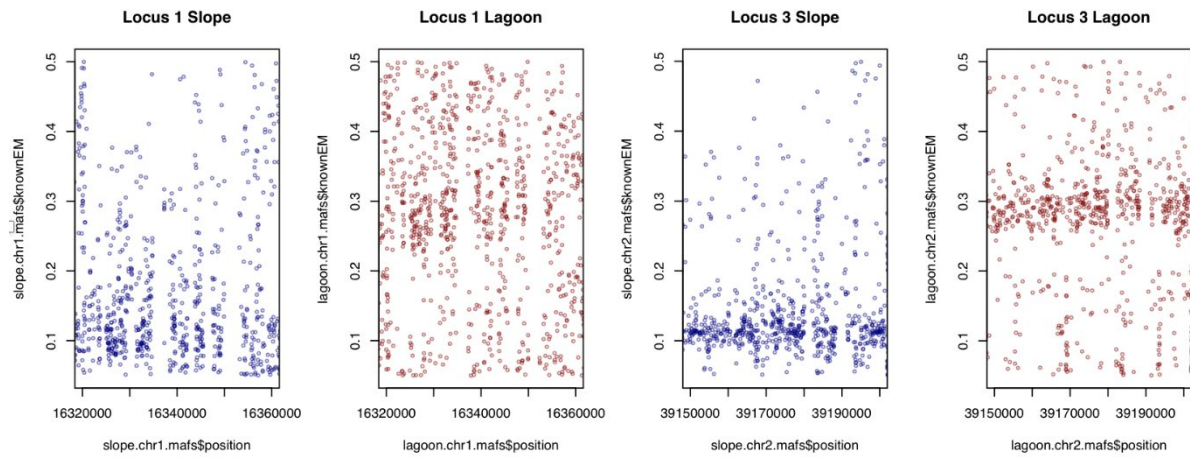


fig. S7 Minor allele frequencies in lagoon and slope samples from Rowley Shoals across Locus 1 and Locus 3.



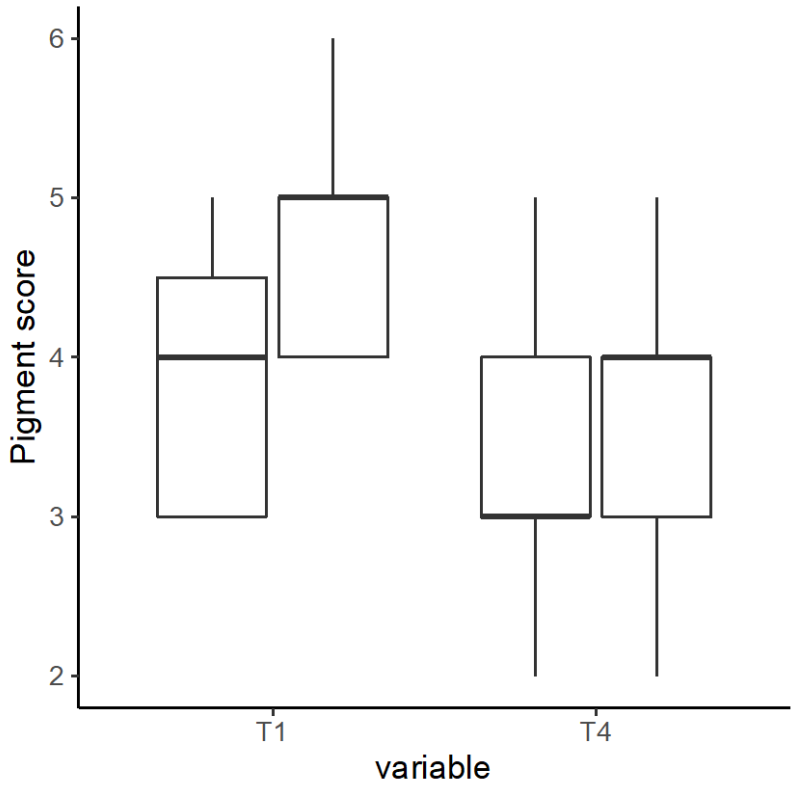


fig. S8 Visual scoring of colonies using the CoralWatch Coral Health Chart from the lagoon (red) and slope (blue) following acute heat stress.

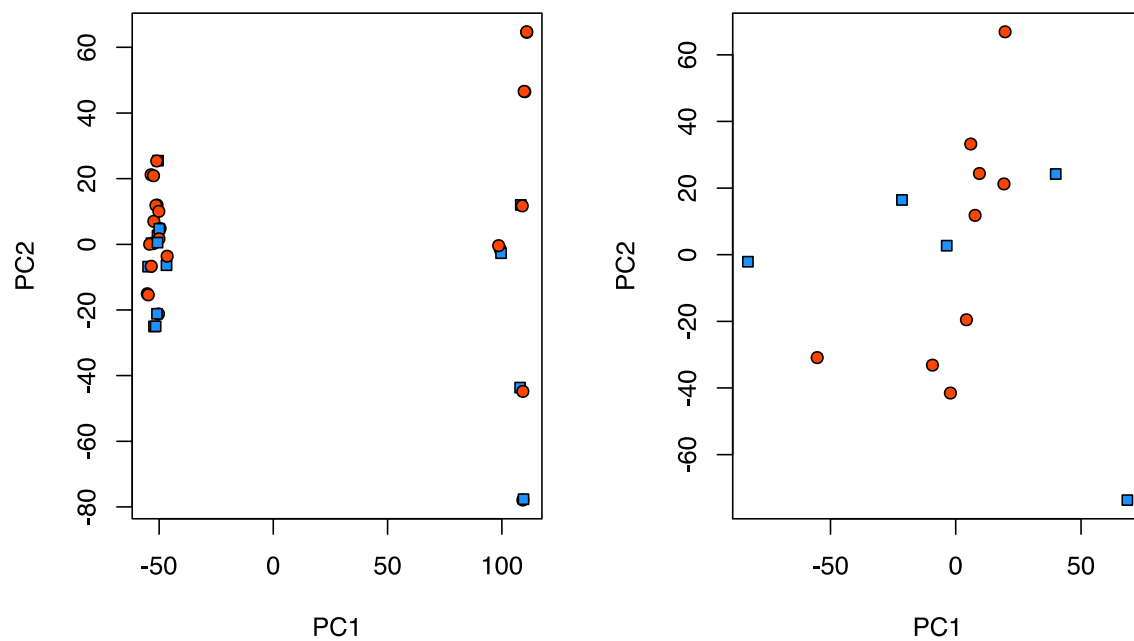


fig. S9 Initial screening of genotypes from RNAseq experiment. Scatter plot of the first two principal components for all genotypes used in experimental heat stress tests (left) and for those after the autumn spawning lineage was removed (right). Colour indicates habitat or origin (red-lagoon, blue-slope). After outlier samples were removed, the RNA sequencing dataset confirmed the negligible divergence (mean  $F_{ST} = -0.004$ ) between habitats that was observed with the low-coverage WGS dataset (fig. S9).

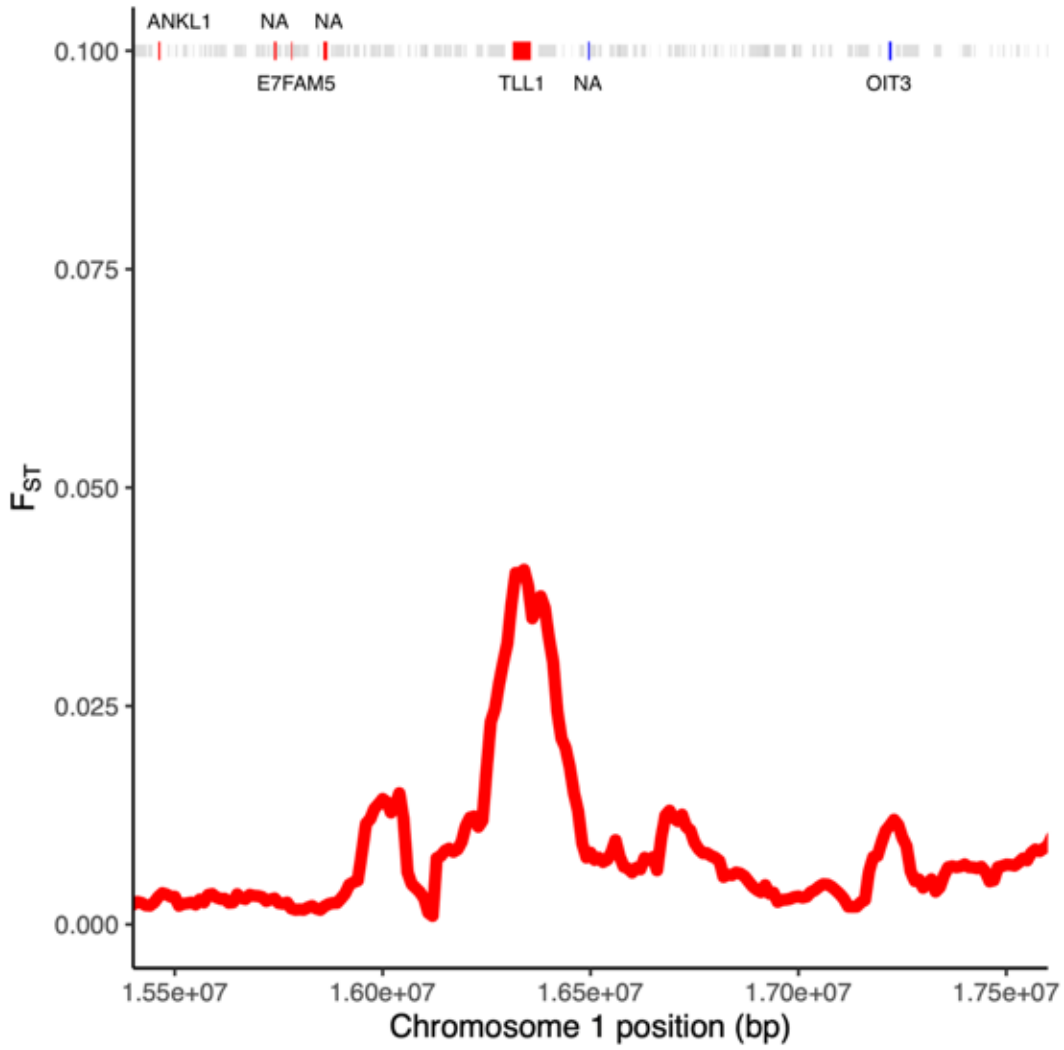


fig. S10 Plot of  $F_{ST}$  for each SNP and in 100kb windows (red line) across Locus 1. Genic regions are highlighted in grey, and ones that were differentially expressed between heated and control colonies are in red and between habitats in blue.

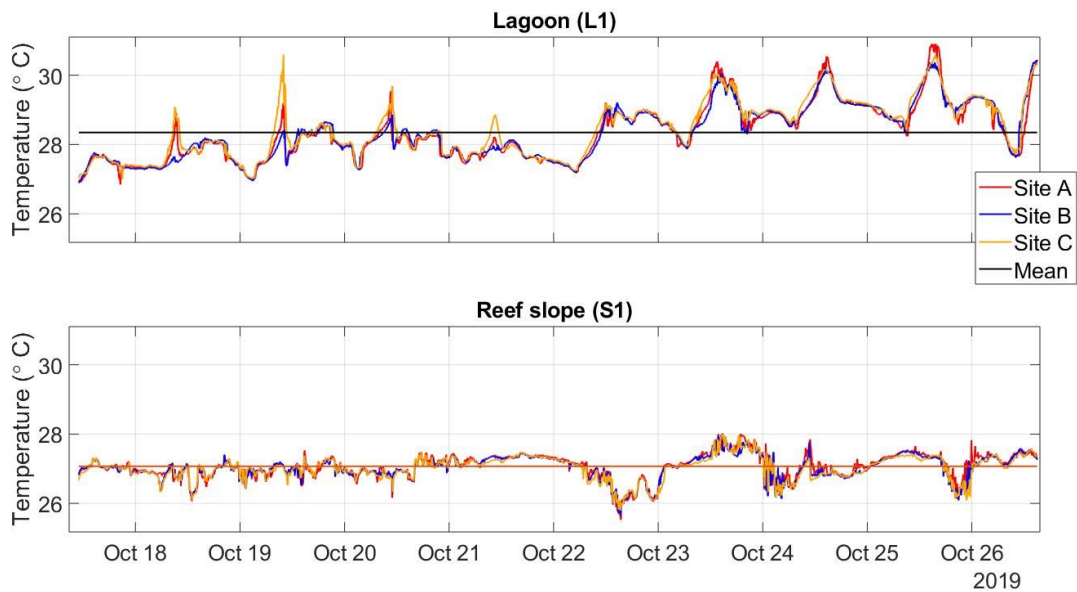


fig. S11 Lagoon and reef slope temperature at sites A, B and C, Clerke Reef. An array of temperature loggers ( $\pm 0.2$  °C accuracy, HOBO U22 Pro-v2) was deployed for 10 days from 17 to 26 October 2019, with 10-min logging intervals at sites within the lagoon (L1) and on the reef slope (S1) of Clerke Reef. The temperature loggers were calibrated in a room-temperature water bath before and after the deployment, ensuring they read to within 0.1°C of each other. The instruments were deployed on top of small lead weights (sampling  $\sim 0.1$ m above the bed) at various water depths depending on the site.

table S10 Mean and temperature variance over the 10-day experiment (17 to 26 October 2019) and mean daily variation at the lagoon site (L1) and reef slope site (S1) for each sub-site.

	<b>Site</b>	<b>Mean (°C)</b>	<b>Variance (°C)</b>	<b>Mean daily variation (°C)</b>
Lagoon (L1)	Site A	28.4	0.6	2
	Site B	28.4	0.6	1.7
	Site C	28.4	0.6	2
Reef slope (S1)	Site A	27	0.1	1.1
	Site B	27	0.1	1.1
	Site C	27	0.1	1

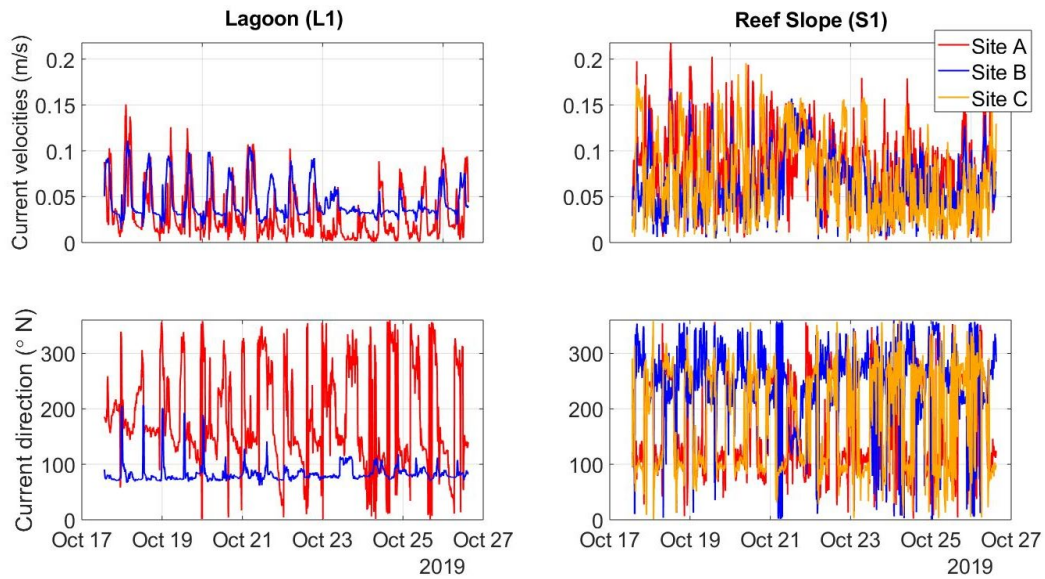


fig. S12 Lagoon and reef slope current velocities and directions at sites A, B and C, Clerke Reef. An array of current meters (Marotte High Sampling Rate (HS) drag-tilt current meters) was deployed for 10 days from 17 to 26 October 2019, with 10-min logging intervals at sites within the lagoon (L1) and on the reef slope (S1) of Clerke Reef. The instruments were deployed on top of small lead weights at various water depths depending on the site.

table S11 Mean and current velocity variance over the 10-day experiment at the lagoon site (L1) and reef slope site (S1) for each sub-site.

	Site	Mean (m/s)	Variance (m/s)
Lagoon (L1)	Site B	0.03	$7 \cdot 10^{-4}$
	Site C	0.04	$3 \cdot 10^{-4}$
Slope (S1)	Site A	0.08	0.002
	Site B	0.06	0.001
	Site C	0.06	0.002

## Cluster Scripts

### 1. Mapping

```
#!/bin/bash
CHUNK=$2
COUNTER=0
FQ="@:3"

for i in $FQ; do
  if [ $COUNTER -eq 0 ]; then
    echo -e "#!/bin/bash\n#SBATCH --ntasks=1\n#SBATCH --cpus-per-task=3\n#SBATCH \
      -t 24:00:00\n#SBATCH --mem 24000" > TEMPBATCH.sbatch; fi
    BASE=$( basename $i _R1.fastq.gz )
    echo "module load bwa" >> TEMPBATCH.sbatch
    echo "bwa mem $1 ${BASE}_R1.fastq.gz ${BASE}_R2.fastq.gz > ${BASE}.sam" >> TEMPBATCH.sbat
ch
    echo "module load samtools" >> TEMPBATCH.sbatch
    echo "samtools view -bSq 10 ${BASE}.sam > ${BASE}_BTVS-UNSORTED.bam" >> TEMPBATCH.sbatc
h
    echo "samtools sort ${BASE}_BTVS-UNSORTED.bam > ${BASE}_UNDEDUP.bam" >> TEMPBATCH.sb
atch
    echo "module load java" >> TEMPBATCH.sbatch
    echo "java -Xmx4g -jar $MYGROUP/software/picard.jar MarkDuplicates REMOVE_DUPLICATES=tru
e \
      INPUT=${BASE}_UNDEDUP.bam OUTPUT=${BASE}.bam METRICS_FILE=${BASE}-metrics.txt \
      VALIDATION_STRINGENCY=LENIENT" >> TEMPBATCH.sbatch
    echo "samtools index ${BASE}.bam" >> TEMPBATCH.sbatch
    echo "rm ${BASE}.sam" >> TEMPBATCH.sbatch
    echo "rm ${BASE}_BTVS-UNSORTED.bam" >> TEMPBATCH.sbatch
    echo "rm ${BASE}_UNDEDUP.bam" >> TEMPBATCH.sbatch
    let COUNTER=COUNTER+1
    if [ $COUNTER -eq $CHUNK ]; then
      sbatch TEMPBATCH.sbatch
      COUNTER=0; fi
  done
  if [ $COUNTER -ne 0 ]; then
    sbatch TEMPBATCH.sbatch; fi
```

### 2. Sequencing coverage

```
#!/bin/bash
samtools depth -a -f $1.bamlist > $1.coverage.txt
awk '{sum=0; for (i=3; i<=NF; i++) { sum+= $i } print sum}' all.coverage.txt | \
awk '{ sum += $1 } END { if (NR > 0) print sum / NR }' > $1.mean.txt
```

### 3. Site frequency spectrum and diversity metrics (unfiltered SFS)

```
#!/bin/bash
export OMP_NUM_THREADS=28

/angsd -bam $1.bamlist -ref aten.chr.fasta -anc aten.chr.fasta -out $1.unfiltered \
  -uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 0 -C 50 -baq 1 \
  -minMapQ 30 -minQ 20 -minInd $2 -setMinDepth $3 -setMaxDepth $4 \
  -GL 1 -doSaf 1 -doCounts 1 -P 16
```

```
/realSFS $1.unfiltered.saf.idx -fold 1 > $1.unfiltered.sfs
```

```
/realSFS saf2theta $1.unfiltered.saf.idx -sfs $1.unfiltered.sfs \  
-outname $1.unfiltered.out
```

```
$MYGROUP/software/angsd/misc/thetaStat do_stat $1.unfiltered.out.thetas.idx \  
-win 1000 -step 1000 -type 0 -outnames $1.slidingwindow.theta
```

#### 4. FST and sliding window

```
#!/bin/bash
```

```
realSFS fst index slope.saf.idx lagoon.saf.idx \  
-sfs slope.lagoon.ml -fstout fst.slope.lagoon -fold 1  
realSFS fst stats fst.slope.lagoon.fst.idx  
realSFS fst stats2 fst.slope.lagoon.fst.idx \  
-win 100000 -step 10000 -type 0 > slidingwindow.slope.lagoon.txt
```

#### 5. Genotype likelihoods with SNP filters

```
#!/bin/bash
```

```
export OMP_NUM_THREADS=28  
srun --export=all -n 1 -c 28 $MYGROUP/software/angsd/angsd -bam $1.bamlist \  
-ref aten.chr.fasta -anc aten.chr.fasta -out $1 \  
-uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -skipTriallelic 1 \  
-minMapQ 30 -minQ 20 -minInd $2 -minMaf 0.05 -SNP_pval 1e-6 -setMinDepth $3 -setMaxDepth  
$4 \  
-GL 1 -doMaf 1 -doMajorMinor 1 -doSaf 1 -doCounts 1 -doGlf 2 -P 28
```

#### 6. Principal Component Analysis

```
#!/usr/bin/env python
```

```
pcangsd.py -beagle $1 -minMaf 0.05 -admix -o $1.pca.admix.outfile -threads 16 -selection
```

#### 7. Linkage disequilibrium

```
#!/bin/bash
```

```
module load gcc  
module load gsl
```

```
## Prepare a pos file by subsampling one SNP in every 10 SNPs in the mafs file
```

```
#zcat < rowleys.beagle.gz | grep "chr1_RaGOO" | awk 'NR % 10 == 0' | cut -f 4- | gzip > rowleys.subs  
ampled.chr1.beagle.gz
```

```
## Prepare a pos file by subsampling one SNP in every 50 SNPs in the mafs file
```

```
#zcat < rowleys.mafs.gz | grep "chr1_RaGOO" | cut -f 1,2 | awk 'NR % 10 == 0' | sed 's:/_g/' | gzip >  
rowleys.subsampled.chr1.pos.gz
```

```
ngsLD \  
--geno rowleys.chr2.ngsLD.beagle.gz \  
--pos rowleys.chr2.ngsLD.pos.gz \  
--probs \  
--min_maf 0.05 \  
--n_ind 72 \  
--n_sites $1 \  
--ignore_miss_data \  
--max_kb_dist 1000 \  
--n_threads 16 \  
--
```

```
--out rowleys.ngsLD.ld
```

```
module load r
```

```
cat rowleys.ngsLD.ld | bash LD_blocks.sh $2 $3 $4
```

## 8. Mitochondrial genome consensus sequence

```
#!/bin/bash
```

```
module load bcftools
```

```
module load samtools
```

```
for i in *.bam
```

```
do
```

```
bcftools mpileup -Ou -f aten_mito.fasta $i | bcftools call -mv -Oz -o $i.vcf.gz
```

```
bcftools index $i.vcf.gz
```

```
cat aten_mito.fasta | bcftools consensus $i.vcf.gz > $i.consensus.fa
```

```
done
```

## 9. Mapping RNAseq reads

```
#!/bin/bash
```

```
CHUNK=$2
```

```
COUNTER=0
```

```
FQ="${@:3}"
```

```
for i in $FQ; do
```

```
  if [ $COUNTER -eq 0 ]; then
```

```
    echo -e "#!/bin/bash\n#SBATCH --partition=day\n#SBATCH --cpus-per-task=10\n#SBATCH --mem=24000" > TEMPBATCH.sbatch; fi
```

```
    BASE=$( basename $i _R1_QC.fastq.gz )
```

```
    echo "/group/oi002/lthomas/software/hisat2-2.2.1/hisat2 -p 3 -X 1500 --rg-id $BASE --score-min L,-0.6,-0.6 --rg SM:$BASE -x $1 -1 ${BASE}_R1_001.fastq.gz -2 ${BASE}_R2_001.fastq.gz > $BASE.sam"
```

```
>> TEMPBATCH.sbatch
```

```
    echo "/group/oi002/lthomas/software/samtools-1.12/samtools view -bSq 10 ${BASE}.sam > ${BASE}_BTVS-UNSORTED.bam" >> TEMPBATCH.sbatch
```

```
    echo "/group/oi002/lthomas/software/samtools-1.12/samtools sort ${BASE}_BTVS-UNSORTED.bam > ${BASE}.bam" >> TEMPBATCH.sbatch
```

```
    echo "/group/oi002/lthomas/software/samtools-1.12/samtools index ${BASE}.bam" >> TEMPBATCH.sbatch
```

```
    echo "rm ${BASE}.sam" >> TEMPBATCH.sbatch
```

```
    echo "rm ${BASE}_BTVS-UNSORTED.bam" >> TEMPBATCH.sbatch
```

```
    let COUNTER=COUNTER+1
```

```
    if [ $COUNTER -eq $CHUNK ]; then
```

```
      sbatch TEMPBATCH.sbatch
```

```
      COUNTER=0; fi
```

```
done
```

```
if [ $COUNTER -ne 0 ]; then
```

```
  sbatch TEMPBATCH.sbatch; fi
```

```
done
```

## 10. Calling variants with samtools

```
#!/bin/bash
```

```
bcftools mpileup -Ou -f aten_0.11.maker_post_001.transcripts.fasta -b all.bamlist -q 20 -Q 20 | bcftools call --skip-variants indels -mv -Ov -o all_snps.vcf
```



```
vcftools --vcf all_snps.vcf --max-missing 0.95 --maf 0.05 --recode --recode-INFO-all --out all_snps.filtered
```

#### 11. Generate counts matrix

```
#!/bin/bash
FIRST=1
for i in "$@"; do
  if [ "$FIRST" -eq 1 ]
  then
    echo "CONTIG" > TEMP-COUNTS-0
    samtools idxstats $i | cut -f 1 >> TEMP-COUNTS-0
    FIRST=0
  fi
  echo $i
  BASE=$(samtools view -H $i | grep @RG | grep -o SM:.* | sed 's/SM://g')
  echo $BASE > TEMP-COUNTS-$i
  samtools idxstats $i | cut -f 3 >> TEMP-COUNTS-$i
done
paste TEMP-COUNTS* > merged_counts.txt
rm TEMP-COUNTS*
```

(Table S1-S9 provided as separate .csv files)

table S1 Microsatellite genotypes to confirm spring lineage.

table S2 Mapping metrics using BWA to *A. tenuis* assembly.

table S3 List of genes with protein BLAST annotation for outlier Locus 1.

table S4 List of genes with protein BLAST annotation for outlier Locus 3.

table S5 Normalized gene expression counts matrix for heated and control samples.

table S6 Results from differential gene expression analyses with DeSeq2 between colonies collected from lagoon and slope habitats after two days acclimation in a common garden.

table S7 Results from differential gene expression analyses with DeSeq2 between heated and control treatments across all samples.

table S8 List of genes differentially expressed between habitats in control treatments that also responded to acute heat stress.

table S9 Counts matrix for the 62 ITS2 sequence variants recovered from colonies of *A. tenuis* from lagoon and slope habitats at Clerke Reef.