

SCIENTIFIC REPORTS

OPEN

Genome-wide SNP analysis explains coral diversity and recovery in the Ryukyu Archipelago

Chuya Shinzato^{1,*}, Sutada Mungpakdee^{1,*}, Nana Arakaki² & Noriyuki Satoh¹

Received: 06 July 2015

Accepted: 12 November 2015

Published: 10 December 2015

Following a global coral bleaching event in 1998, *Acropora* corals surrounding most of Okinawa island (OI) were devastated, although they are now gradually recovering. In contrast, the Kerama Islands (KIs) only 30 km west of OI, have continuously hosted a great variety of healthy corals. Taking advantage of the decoded *Acropora digitifera* genome and using genome-wide SNP analyses, we clarified *Acropora* population structure in the southern Ryukyu Archipelago (sRA). Despite small genetic distances, we identified distinct clusters corresponding to specific island groups, suggesting infrequent long-distance dispersal within the sRA. Although the KIs were believed to supply coral larvae to OI, admixture analyses showed that such dispersal is much more limited than previously realized, indicating independent recovery of OI coral populations and the necessity of local conservation efforts for each region. We detected strong historical migration from the Yaeyama Islands (YIs) to OI, and suggest that the YIs are the original source of OI corals. In addition, migration edges to the KIs suggest that they are a historical sink population in the sRA, resulting in high diversity. This population genomics study provides the highest resolution data to date regarding coral population structure and history.

Coral reefs, which host one-third of all described marine species, are declining due to various anthropogenic insults, resulting in ocean acidification, seawater temperature increases, increased sediment, higher nutrient and pollutant loads, greater frequency and intensity of storms, etc^{1–4}. Understanding coral genetic diversity, connectivity, and population structure is essential to increase the effectiveness of coral restoration efforts^{5–7}. Corals exhibit a typical metapopulation structure with each population comprising many sub-populations connected by larval dispersal that occurs during the planktonic larval phase⁸. When a population is severely damaged, larvae supplied from undisturbed populations are essential for recovery⁹.

The southern Ryukyu Archipelago (sRA) has unique geographic, hydrodynamic, and historical features. It is located in southwestern Japan, spanning over 600 km (Fig. 1), and the strong Kuroshio ocean current runs from the eastern Philippines through the Archipelago (Fig. 1b). Coral reefs throughout the sRA are regarded as a marine biodiversity hotspot¹⁰. In 1998, a large percentage of *Acropora* corals at OI were lost as a result of a global coral bleaching event^{11,12}, but they are now gradually recovering^{13,14}. In contrast, the Kerama Islands (KIs), located approximately 30 km west of OI have retained healthy corals and high diversity, suffering only 6.7 to 23.4% mortality around Aka Island during the mass bleaching¹⁵, and maintaining high coral coverage (23–35%) from 2003 to 2007¹⁴. The Keramas became a Japanese National Park in 2014. Given the stark differences in the recent history of these geographically close populations, we set out to answer several questions. How have corals recovered at OI? How were the KIs able to retain such diverse, healthy coral populations? To what extent did KI corals contribute to the OI coral recovery? Next-generation sequencing (NGS) has allowed us to sample genomes much more densely and to observe the patterns of genetic variation that result from the full range of evolutionary processes acting across the genome¹⁶. Now it is possible to identify thousands of single nucleotide polymorphisms (SNPs) throughout a genome, and to acquire vast quantities of genomic information from a sample, whereas formerly it was only possible to examine small numbers of neutral molecular markers (e.g. allozymes, microsatellites) covering a very limited sampling of the genome. Today we can improve the precision of population genomic analyses by greatly increasing the number of markers examined. Recently, genome-wide SNP data have become widely used

¹Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan. ²DNA Sequencing Section, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to C.S. (email: c.shinzato@oist.jp)

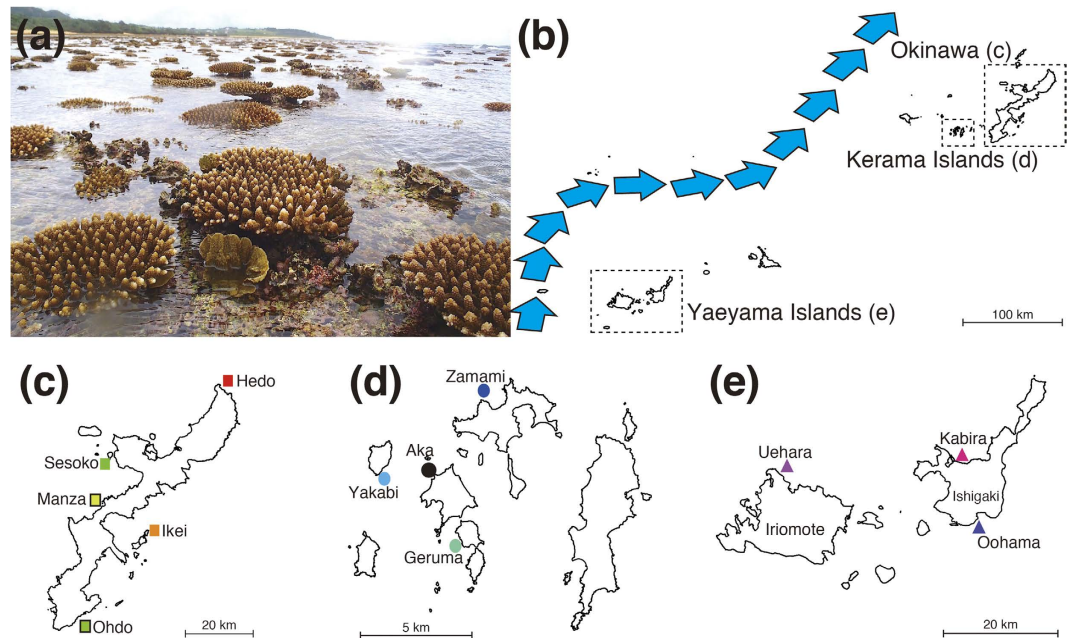


Figure 1. A scleractinian coral, *Acropora digitifera*, and sampling locations in the southern Ryukyu Archipelago. (a) Typical habitat of *A. digitifera* colonies used in this study. Some colonies are exposed to air at low tide. Photo taken at Uehara, Iriomote Island. (b) Overview of Okinawa prefecture. Blue arrows indicate the Kuroshio Current in May 2014, based on the Japan Meteorological Agency website (<http://www.jma.go.jp/jma/indexe.html>). (c) Sampling sites at Okinawa Island, (d) the Kerama islands, (e) the Yaeyama islands (Ishigaki and Iriomote islands are identified). A shape file for Okinawa prefecture was downloaded from National Land Numerical Information, Japan (<http://nlftp.mlit.go.jp/ksj/gmlold/index.html>) and was visualized as maps using maptools in R (version 3.1.1)³¹. All maps are oriented with North at the top.

for studying population structure and historical migration patterns of various organisms^{17,18} including fishes^{19,20} and a seahorse²¹. Thus NGS technologies have become powerful tools in conservation ecology¹⁶.

The stony coral, *Acropora digitifera*, is widely distributed in the Indo-Pacific Ocean (Fig. 1a) and it is the only coral species with a decoded genome²². *A. digitifera* is a broadcast-spawning coral and typically inhabits the shallow, flat reef zone in Okinawa (Fig. 1a). Waves and ocean currents should significantly influence larval dispersal. Larval culture experiments with *A. digitifera* have shown that the maximum settlement competency period of larvae, the period within which larvae can settle on a suitable substrate, is 45 days²³. An investigation of ocean current flow between the KIs and OI using high-frequency radar and GPS buoys showed that coral larval could drift from the KIs to OI in 3 to 4 days, implying that the KIs are an important larval source for the west-coast of OI²⁴. Moreover, a study of *A. digitifera* population genetics using microsatellite markers suggested high genetic connectivity across the entire RA²⁵. However a detailed understanding of population structure was nonexistent. Using the decoded genome of *A. digitifera*, we performed the first genome-wide SNP analysis of corals and succeeded in mapping detailed population structure and migration patterns of *A. digitifera* in the sRA.

Results and Discussion

We first improved the original *A. digitifera* genome assembly²². The new version (1.1) is approximately 447 Mbp and comprises 2,420 scaffolds with an N50 of 484 Kbp (Supplementary Table S1). Longer-range continuity of the genome assembly and an increase in the N50 scaffold size from 191 kbp to 484 kbp (Supplementary Table S1), allowed us to assess long-range linkage disequilibrium. In addition, the total nucleotide count without ambiguous bases (gaps) increased from 365 Mbp to 378 Mbp, indicating that the new reference genome could produce more SNP data than the previous version. Thus we concluded that the quality is significantly improved compared with the original release²².

We collected a total of 155 individual colonies from OI, the KIs and the Yaeyama Islands (YIs) (Fig. 1b–e). Genomes of all samples were re-sequenced and SNPs of each sample were identified. We excluded individuals with heterozygosity rates >0.12 and/or missing genotype rates >0.35 (Supplementary Fig. 1). Subsequent removal of SNPs that showed extensive deviation ($p < 0.001$) from Hardy-Weinberg equilibrium (HWE), with a call rate less than 99% and with a very low minor allele frequency (threshold of 2%), yielded 905,561 SNPs. After removal of related sample outliers, 122 individuals from 11 locations remained (all individuals from the Aka site were removed), and these were used for subsequent analyses.

Linkage disequilibrium decayed within <10 kbp at all sampling locations and we could not observe different decay patterns between sampling locations (Supplementary Fig. 2), indicating no population-specific bottlenecks in these areas. We tried to estimate individual ancestries of *A. digitifera* populations in the sRA using model-based clustering methods, ADMIXTURE²⁶ and fastSTRUCTURE²⁷ (for an example, see Supplementary Fig. 3), but we

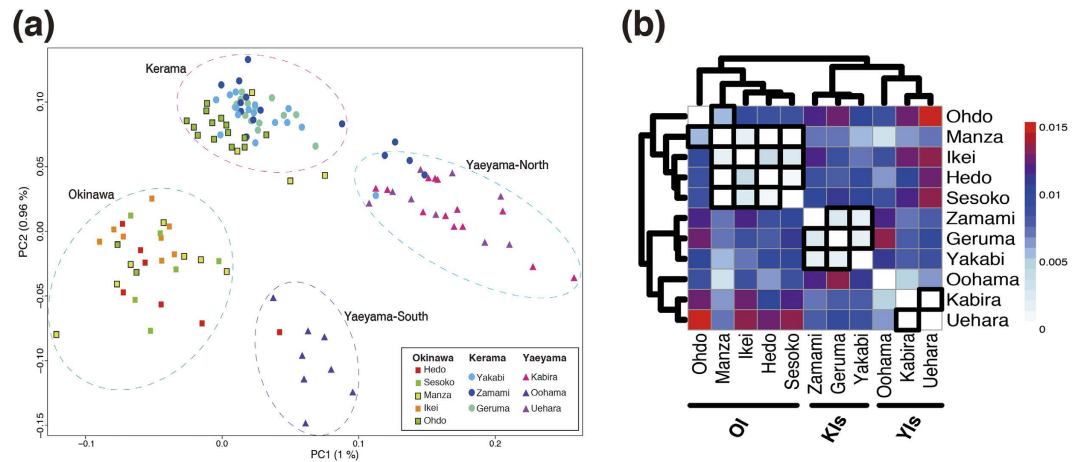


Figure 2. Principal components analysis and genetic distances suggest the existence of four clusters in the sRA, located in Okinawa, the KIs, Yaeyama North, and Yaeyama South. (a) Principal components analysis (PCA) of the filtered dataset of 122 individuals and 905,561 SNPs after trimming of related samples with smartrel and outliers with smartpca. Tracy-Widom statistical significance of PC1 and PC2 was $< 1e^{-35}$. Individuals from Okinawa Island, the KIs, and the YIs are shown with squares, circles, and triangles, respectively. Four subpopulations, Okinawa, Kerama, Yaeyama-North, and Yaeyama-South are circled in green, pink, blue, and purple dotted lines, respectively. (b) Heatmap showing pairwise *Fst* values based on Weir and Cockerham weighted estimates between sampling sites. All points were clustered by pairwise *Fst* values, based on the Complete Linkage Clustering method. Site combinations without boxed bold lines indicate significant differences (p-value < 0.001, ANOVA statistics across 10 eigenvectors in PCA).

failed to detect clear population structure. In both software packages, the appropriate number of populations (*K*) to best explain acroporid genetic differentiation was *K* = 1 (*K* can range from 1 to 10), suggesting that there is a single population of *A. digitifera* in the sRA.

Although model-based clustering methods for estimating ancestry could not detect clear population structure, principle component analysis (PCA) resolved the 122 individuals into four clusters, such that individuals from the same island or close locations tended to fall into the same cluster. Thus we named these Okinawa, Kerama, Yaeyama-North, and Yaeyama-South, respectively (Fig. 2a). The first two principal components explained 1% and 0.96%, respectively, of the total variance among SNP data of 122 individuals. Two distinctive clusters in YI possibly originated from larval dispersal interference caused by Sekisei lagoon, the largest coral reefs in Japan, located between Ishigaki and Iriomote islands. Most individuals from the KIs (Yakabi, Geruma and Zamami) were clustered together. Individuals from three points in Okinawa (Oku, Ikei, and Sesoko) were observed exclusively in the Okinawa cluster (Fig. 2a). On the other hand, a large number of individuals from Ohdo and some individuals from Manza were observed in both the Kerama and Okinawa clusters (Fig. 2a). This suggests that recent migration from the KIs to OI (mainly south coast, some west coast) has occurred and that these might be first-generation migrants. In addition, some individuals in the KIs (mainly Zamami) had SNP profiles similar to those of the Yaeyama-North cluster, and an individual from Hedo (the northern tip of Okinawa) was grouped with the Yaeyama-South cluster (Fig. 2a), suggesting recent migration events between the KIs and Yaeyama-North and Okinawa and Yaeyama-South, respectively.

Slight genetic differentiation estimated by pairwise *Fst* values for all site combinations (−0.0007 to 0.0138) also suggested that there is a single population in the region. Although low *Fst* values were estimated for all combinations, sampling locations on the same island or close locations were clustered together (Fig. 2b). Previous coral population genetics studies using allozymes and microsatellites reported no correlation between genetic connectivity and geographic distance in *A. digitifera* in the sRA^{23,25}. However we detected a significant isolation-by-distance correlation at the geographic scale of the study ($P < 0.001$ and $R^2 = 0.3715$; Supplementary Fig. 4). ANOVA statistics across 10 eigenvectors in PCA (Fig. 2a) also revealed that sampling locations that were not significantly different ($P > 0.001$) are clustered on the same island or in proximate locations (Fig. 2b), as in the case of *Fst* values. Previous studies suggested that high genetic connectivity exists across the 1,000 km length of the RA²⁵ and across the 1,500 km distance between the RA and the Ogasawara Islands. Those results implied that long-distance dispersal of *Acropora* larvae between high-latitude islands in Japan²⁸ might be frequent; however, the larger dataset in this study makes it clear that this is not the case.

This is neither surprising, nor is it an isolated example. In *Seriatopora hystrix*, as in other brooding corals, the majority of larval recruitment in northern Western Australia occurred within 100 m of the natal colony²⁹ and long-distance dispersal among high-latitude reefs in eastern Australia rarely occurred³⁰. Even in cases of broadcast spawners, it is reported that two coral species possessing contrasting reproductive modes (*Goniastrea favulus*: sticky, negatively buoyant eggs and larvae, *Platygyra daedalea*: positively buoyant egg-sperm bundles) are genetically subdivided by distances as small as 50–100 m in the southern Great Barrier Reef³¹. A majority of *Montipora capitata* colonies are derived from self-recruitment across the Hawaiian Archipelago³². Although

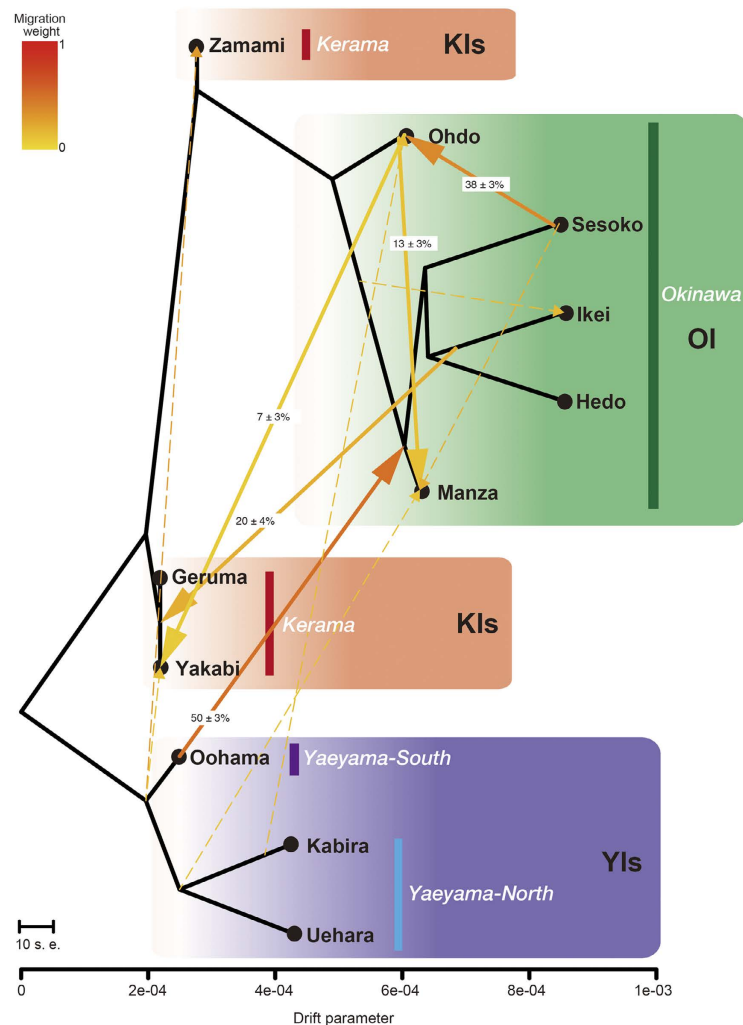


Figure 3. Migration patterns of *A. digitifera* in the sRA, showing that the YI populations function as source populations for OI, while the KIs constitute a historical sink. Inferred tree of *A. digitifera* populations in the sRA with thirteen migration events. Migration arrows are colored according to their weight. The migration weight represents the fraction of ancestry derived from the migration edge. Migration edges with p-value estimated by jackknife to be below 0.01 are shown. Migration edges with bootstrap support less than 50% are shown with dotted lines. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred in each branch.

local biological, ecological, geographic factors, and ocean currents could have great influence on larval dispersal and coral recruitment, long-distance migration may be less frequent than suggested and may not occur in every generation. Instead, self-recruitment may be the major source of larval recruitment of *Acropora* corals in the sRA and may play an important role in coral recovery recently observed in OI¹⁴.

How were the coral reefs established, and how have population structures been maintained? We constructed admixture graphs using *TreeMix* to examine the population history of *A. digitifera* in the sRA. The maximum-likelihood population tree inferred without admixture events was similar to the clustering result obtained with *Fst* values (Fig. 2b), clearly correlating genetics with island locations (Supplementary Fig. 5a). Next we sequentially allowed for 0 to 15 migration events in the *TreeMix* analysis (Fig. 3). The increase in likelihood beyond thirteen such events was marginal, and stepwise comparisons of log likelihood between migration events became insignificant between 13 and 14 events (likelihood-ratio test, $p > 0.05$) (Supplementary Fig. 6). Thus we chose thirteen migration events for the *TreeMix* analysis. Five migration edges had more than 50% bootstrap support (Fig. 3, Supplementary Table 4). We detected strong long-distance migration spanning 400 km, from the YIs to OI. Individuals from Oohama, Ishigaki Island, and Hedo, Sesoko, Ikei, and Manza in Okinawa share $50 \pm 3\%$ ancestry (Fig. 3). Three-population tests also supported such migration (Supplementary Table 3). These suggest that the *A. digitifera* population in the sRA is partially maintained by long-distance migrations across the sRA; thus low genetic distance (*Fst* values, Fig. 2b) and common ancestry estimated by model-based clustering methods are observed in this study (Supplementary Fig. 3).

Complex migration patterns within OI were further inferred: migrations from Sesoko to Ohdo, and Ohdo to Manza (Fig. 3), and migrations from OI to the KIs were also detected (Ikei to Geruma and Yakabi, and Ohdo to Yakabi, Fig. 3). Three-population tests indicated active migrations from a variety of locations to Yakabi Island (KIs) (Supplementary Table 3). Although no individuals from Ikei were detected in the Kerama cluster using PCA (Fig. 2a), a migration from Ikei to the KIs was detected (Fig. 3, Supplementary Table 3). This suggests that coral connectivity from the east coast of OI to the KIs is possible. Although bootstrap support is poor (<50%, Supplementary Table 4), we detected migrations from the YIs to the KIs (Fig. 3). Heterozygosity of the KIs, especially Geruma and Yakabi, are significantly higher than those of OI and the YIs (Supplementary Fig. 7). Although maintenance of high coral coverage in the KIs¹⁴ would explain the high heterozygosity in that area, the heterozygosity might also reflect past migrations from both OI and the YIs to the KIs. Low heterozygosity of OI and the YIs might reflect a recent population bottleneck caused by a significant decrease of *Acropora* coral coverage after the 1998 bleaching event^{11,12,33}.

Previous research using high-frequency radar and global positioning system buoys to study the length of the *Acropora* larval settlement competency period, suggested that coral larval transport had occurred from the KIs to OI²⁴. In 1998, a massive bleaching event occurred at OI¹¹ and after a decade, *Acropora* colonies slowly recovered along the west coast of Okinawa¹⁴, including Sesoko¹³. Van Woesik *et al.* (2011) proposed that the KIs facilitated *Acropora* recovery on Sesoko by supplying recruits. However the present PCA analysis did not suggest that any Sesoko individuals came from the Kerama cluster (Fig. 2a) and no migration from Kerama to Sesoko was supported by *TreeMix* or the three-population analysis (Supplementary Table 3). Thus it is improbable that the KIs were a major source for *Acropora* recovery in OI. A migration to the south coast of OI (Ohdo) was suggested by PCA (Fig. 2a); however, it is apparent that the KIs are a far more limited larval source than previously thought. Instead it seems that coral recovery in Sesoko occurred via recruitment from local sources that survived the bleaching event, or that it regrew from cryptic remnant tissues^{34,35}. Conversely our results of *TreeMix* and three-population test showed that the YI populations have a strong influence on OI populations (Fig. 3). Recent long-distance migration, especially from Ishigaki to OI, might explain the low genetic distances (*Fst*) (Fig. 2b) detected in this study.

In conclusion, this first coral population genomics study, analyzing genome-wide SNP data, has considerably illuminated the recent history of acroporid coral population structure in the sRA, even though genetic distances are quite low in the area. (1) Coral recovery observed in the last decade in OI has been accomplished mainly by self-recruitment. Because of low gene flow between locations, local conservation and restoration efforts will be required for each coral reef in the sRA. (2) Migrations from both OI and the YIs to the KIs were documented and suggest that the KIs are a “historical sink” of coral connectivity in this region, maintaining high coral diversity. (3) The YI populations have historically influenced the establishment and maintenance of OI populations. These findings, resulting from the high resolution of next-generation sequencing technology, provide data that will be essential for future coral reef conservation in the region.

Although re-sequencing-based genome-wide SNP analyses require reference genomes and are restricted to species with decoded genomes, now reference-free methods for identifying genome-wide SNPs are available (e.g. RAD-seq³⁶). Application of RAD-seq to coral population genetics³⁷ and linkage mapping³⁸ have been reported; however, the number of SNPs identified was much smaller than by the whole genome re-sequencing method used in this study. Re-sequencing-based, genome-wide SNP data enable us to obtain large amounts of biological information, and will become an important and powerful tool for coral biology.

Methods

Sample collection. Tissue samples were collected from 155 individual *Acropora digitifera* colonies in the southern RA from 2011 to 2014 under Okinawa Prefectural Permits (Numbers: 22–29, 23–34, 24–48, 25–67). Collection sites include five locations (Hedo, Sesoko, Manza, Ikei, and Ohdo) on Okinawa Island, four locations (Geruma, Yakabi, Zamami, and Aka) in the Kerama Islands, two locations (Oohama and Kabira) near Ishigaki Island, and one (Uehara) at Iriomote Island in the Yaeyama Islands (Fig. 1). We identified *A. digitifera* in the reefs and used colonies showing typical corymbose shape (Fig. 1a). As *A. digitifera* corals in the RA are often found around reef-edges, most samples were collected at depths of 1–2 m, where some corals are exposed to the air during low tide (Fig. 1a). All Kerama samples came from slightly deeper reefs (1–4 m). To avoid duplicate collections of colonies that could have been produced through asexual fragmentation or propagation, only colonies that were physically distinct and at least 5 m from other colonies were sampled. A single branch about 1–1.5 cm long was taken from each colony using a nipper. Samples were preserved in the guanidinium reagent, CHAOS³⁹, and stored at room temperature.

DNA isolation and genome re-sequencing. Coral branches in CHAOS buffer were incubated at 56 °C for 48 hours. After that, coral skeletons were removed and preserved in the event that future species confirmation using skeletal characters might be needed. After ethanol precipitation, pellets were dissolved in ALT buffer from a QIAGEN DNeasy blood and tissue kit (cat. 69504) and then they were digested with proteinase K for 48 hours at 56 °C. Samples were further purified according to QIAGEN kit instructions, including RNaseA treatment. At the end, DNAs were eluted from spin columns using AE buffer and kept at 4 °C. Library preparation for re-sequencing of DNA from each individual genome using Illumina platforms (500 bp insert size) was done according to the manufacturer's instructions. Libraries were sequenced using an Illumina Genome Analyzer IIx (GAIIx) (75 bp paired-end reads) and HiSeq 2000 (100 bp paired-end reads). Raw sequence data were submitted to the DNA Databank of Japan (DDBJ) Sequence Read Archive (DRA) under accession number DRA003938 (BioProject PRJDB4188).

SNP identification and detection. To improve assembly of the reference genome of *A. digitifera*, we used sequence data from an individual from Hedo, Okinawa, reported in Shinzato *et al.* (2011)²², and further scaffolding was performed with SSPACE (ver. 1.1)⁴⁰. Gaps inside the scaffolds were closed using GapCloser (ver. 1.10)⁴¹. To overcome potential assembly errors arising from tandem repeats, sequences that were aligned to another sequence over 50% of their lengths with BLASTN (e-value: $1e^{-50}$) were removed from the assembly. The genome assembly has been deposited with the DDBJ under project accession BACK02000001–BACK02054400 (contigs) and DF970692–DF973111 (scaffolds).

All Illumina reads from each individual were mapped to the reference sequence using BWA version 0.5.9⁴². We applied a soft-clipping base quality threshold of 10 (phred scaled) to avoid low quality bases in alignments, but otherwise used default settings. Unmapped reads and duplicated reads were removed using SAMtools version 0.1.18⁴³. When reads map to multiple locations in the genome, BWA randomly assigns the hit to one location, with a mapping score of zero. Therefore, such reads are taken into account in coverage estimates, but the low mapping score downplays their effect on downstream variant calling. Results for each individual were enhanced by local realignment using the GATK toolkits version 2.7⁴⁴ and duplicate-marked with Picard version 1.6 (<http://picard.sourceforge.net>).

Initial variant site identification was performed using low-coverage, multi-sample SNP calling with SAMtools mpileup and GATK with default settings from the realigned bam file. Consensus calls (~17.8 million SNPs) from both software packages were further filtered by including only no-missing-call SNPs and used as a known site for base quality recalibration using GATK with default settings. The resulting recalibrated bam files were then used as new input for variant calling with GATK. Then the highest scoring 10% of consensus calls (~75,000 SNPs) were used as a training set for variant quality recalibration and filtering with GATK.

Data quality control and markers. Generally we followed the protocol for data quality assessment and control that is typically employed during genome-wide case-control association studies⁴⁵. Samples with low DNA quality or concentration often have below average call rates and genotype accuracy. Therefore, the genotype failure rate and heterozygosity rate per individual are both measures of DNA sample quality. We inspected the distribution of mean heterozygosity across all individuals to identify those with an excessive or reduced proportion of heterozygous genotypes, which could indicate DNA sample contamination, fragmentation, or DNA from a different *Acropora* species. First, missing genotype and heterozygosity rates of each individual were calculated using Plink, version 1.9⁴⁶. Observed heterozygosity rates per individual were calculated using the formula $(N - O)/N$, where N is the number of non-missing genotypes and O the observed number of homozygous genotypes for a given individual⁴⁵. We used the Tukey-Kramer test to check differences in heterozygosity rates among all site combinations. Since the smallest number of samples from any sampling site was six (Sesoko), we used six individuals from each point to calculate pairwise linkage disequilibrium, using no missing SNP data with Plink. To avoid duplicated or related individuals, only independent SNPs were included in the analysis. SNPs exhibiting extended linkage disequilibrium (LD) in which a correlated ($r^2 = 0.2$) pair of SNPs within a window of 50 kb, were removed from the data set using Plink. We excluded markers that showed extensive deviation ($p < 0.001$) from HWE, with a call rate $\leq 99\%$, a missing rate of 1%, and with a very low minor allele frequency (MAF) (threshold of 2%) using Plink. We used this SNP data set for all analyses in this study. Pruned data were subjected to smartrel from EIGENSOFT (version 5.0.2)^{47,48} to identify related samples and smartpca was used to identify outliers, to perform Principal Component Analysis (PCA), and to assess significant differences between sampling points by ANOVA across eigenvectors.

Fst analysis. *Fst* values were calculated for bi-allelic sites using the method of Weir and Cockerham with vcftools version 0.1.12b⁴⁹. *Fst* values were visualized with a heatmap and clustered using the complete linkage method for hierarchical clustering using the R package “pheatmap” (<http://cran.r-project.org/web/packages/pheatmap/index.html>). A plot of isolation-by-distance was used to explore the relationship between the calculated geographic distance and $(1 - Fst)/Fst$, as a genetic differentiation index. The log of coastal distances between sampling locations (km) was determined using Google Earth. IBD significance was tested using Pearson's correlation (R, version 3.1.1)⁵⁰.

TreeMix and admixture analyses. In order to estimate historical relationships among populations, we used TreeMix version 1.12⁵¹. For all analyses we used an SNP window size ($-K$) of 1,000. This corresponds to a window size of approximately 0.46 Mbp (the expected genome size of *A. digitifera* is 420 Mbp and a window size of 1,000 divided 905,561 SNPs into 181 blocks), which far exceeds the known extent of LD (see Supplementary Fig. 2) in *A. digitifera*. TreeMix was run using the $-global$ option which corresponds to performing a round of global rearrangements of the graph after initial fitting. In addition, 100 bootstrap replicates were performed using the $-bootstrap$ option for numbers of migration events ranging from 3–14 ($-m$ option) to assess statistical support for certain migration events. We performed a likelihood-ratio test for stepwise comparison of log likelihood between migration events. The three-population test is a formal test of admixture that provides clear evidence of admixture, even if the gene flow events occurred hundreds of generations ago⁵²; when $f_3(X; A, B)$ is negative, deviation from “treeness” is detected and X appears to be a mixture of A and B . For the three-population test, we used the TreeMix-implemented three-population test⁵³. We used all combinations of three points out of the 11 sampling sites.

References

- De'ath, G., Fabricius, K. E., Sweatman, H. & Puotinen, M. The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proc Natl Acad Sci USA* **109**, 17995–17999, doi: 10.1073/pnas.1208909109 (2012).
- Hoegh-Guldberg, O. *et al.* Coral reefs under rapid climate change and ocean acidification. *Science* **318**, 1737–1742, doi: 10.1126/science.1152509 (2007).
- Jackson, J., Donovan, M., Cramer, K. & Lam, V. *Status and trends of Caribbean coral reefs: 1970–2012*. 306 (Gland, Switzerland, 2014).

4. Knowlton, N. *et al.* In McIntyre A. D. (ed.). *Life in the World's Oceans: Diversity, Distribution, and Abundance* 65–79 (Wiley-Blackwell, 2010).
5. Almany, G. R. *et al.* Connectivity, biodiversity conservation and the design of marine reserve networks for coral reefs. *Coral Reefs* **28**, 339–351, doi: 10.1007/S00338-009-0484-X (2009).
6. Baums, I. B. A restoration genetics guide for coral reef conservation. *Mol Ecol* **17**, 2796–2811, doi: 10.1111/j.1365-294X.2008.03787.x (2008).
7. Shearer, T. L., Porto, I. & Zubillaga, A. L. Restoration of coral populations in light of genetic diversity estimates. *Coral Reefs* **28**, 727–733, doi: 10.1007/s00338-009-0520-x (2009).
8. Bak, R. P. M. & Meesters, E. H. Population structure as a response of coral communities to global change. *Am Zool* **39**, 56–65 (1999).
9. Palumbi, S. R. Population genetics, demographic connectivity, and the design of marine reserves. *Ecol Appl* **13**, S146–S158 (2003).
10. Roberts, C. M. *et al.* Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**, 1280–1284, doi: 10.1126/science.1067728 (2002).
11. Loya, Y. *et al.* Coral bleaching: the winners and the losers. *Ecol Lett* **4**, 122–131 (2001).
12. Hongo, C. & Yamano, H. Species-specific responses of corals to bleaching events on anthropogenically turbid reefs on Okinawa Island, Japan, over a 15-year period (1995–2009). *PLoS One* **8**, e60952, doi: 10.1371/journal.pone.0060952 (2013).
13. van Woesik, R., Sakai, K., Ganase, A. & Loya, Y. Revisiting the winners and the losers a decade after coral bleaching. *Mar Ecol-Prog Ser* **434**, 67–76, doi: 10.3354/Meps09203 (2011).
14. Iwata, K. & Sakai, K. Factors for the Annual Variation of Number of Settled Acroporid Larvae in the West Coast of Okinawa Island, Southeastern Japan. *Biol. Mag. Okinawa* **48**, 1–10 (2010) (in Japanese).
15. Iwao, K. *Study on the effect of geographical features on the cause of coral bleaching*. 15–39 (Research and Development Bureau, Science and Technology Agency Japan (2000) (in Japanese).
16. Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R. & Hohenlohe, P. A. Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol* **22**, 2841–2847, doi: 10.1111/mec.12350 (2013).
17. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413, doi: 10.1038/nature13673 (2014).
18. Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, 1255832, doi: 10.1126/science.1255832 (2014).
19. Hemmer-Hansen, J., Therkildsen, N. O. & Pujolar, J. M. Population Genomics of Marine Fishes: Next-Generation Prospects and Challenges. *Biol Bull-Us* **227**, 117–132 (2014).
20. Karlsen, B. O. *et al.* Genomic divergence between the migratory and stationary ecotypes of Atlantic cod. *Mol Ecol* **22**, 5098–5111, doi: 10.1111/mec.12454 (2013).
21. Boehm, J. T., Waldman, J., Robinson, J. D. & Hickerson, M. J. Population Genomics Reveals Seahorses (*Hippocampus erectus*) of the Western Mid-Atlantic Coast to Be Residents Rather than Vagrants. *PLoS One* **10**, e0116219, doi: 10.1371/journal.pone.0116219 (2015).
22. Shinzato, C. *et al.* Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* **476**, 320–323, doi: 10.1038/nature10249 (2011).
23. Nishikawa, A. & Sakai, K. Settlement-competency period of planulae and genetic differentiation of the scleractinian coral *Acropora digitifera*. *Zoolog Sci* **22**, 391–399 (2005).
24. Nadaoka, K. *et al.* Larval Tracking Using Small Drifters and Larval Settlement Experiments to Examine Long-Distance Larval Transport of Corals. *JSCE* **49**, 366–370 (2002) (in Japanese).
25. Nakajima, Y., Nishikawa, A., Iguchi, A. & Sakai, K. Gene flow and genetic diversity of a broadcast-spawning coral in northern peripheral populations. *PLoS One* **5**, e11149, doi: 10.1371/journal.pone.0011149 (2010).
26. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–1664, doi: 10.1101/gr.094052.109 (2009).
27. Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* **197**, 573–589, doi: 10.1534/genetics.114.164350 (2014).
28. Nakajima, Y., Nishikawa, A., Iguchi, A. & Sakai, K. Regional genetic differentiation among northern high-latitude island populations of a broadcast-spawning coral. *Coral Reefs* **31**, 1125–1133, doi: 10.1007/S00338-012-0932-X (2012).
29. Underwood, J. N., Smith, L. D., Van Oppen, M. J. & Gilmour, J. P. Multiple scales of genetic connectivity in a brooding coral on isolated reefs following catastrophic bleaching. *Mol Ecol* **16**, 771–784, doi: 10.1111/j.1365-294X.2006.03187.x (2007).
30. Noreen, A. M., Harrison, P. L. & Van Oppen, M. J. Genetic diversity and connectivity in a brooding reef coral at the limit of its distribution. *Proc Biol Sci* **276**, 3927–3935, doi: 10.1098/rspb.2009.1050 (2009).
31. Miller, K., J. & Ayre, D. J. Population structure is not a simple function of reproductive mode and larval type: insights from tropical corals. *J Anim Ecol* **77**, 713–724 (2008).
32. Concepcion, G. T., Baums, I. B. & Toonen, R. J. Regional population structure of *Montipora capitata* across the Hawaiian Archipelago. *B Mar Sci* **90**, 257–275, doi: 10.5343/bms.2012.1109 (2014).
33. Kayanne, H. *et al.* Changes in living coral coverage before and after the 1998 bleaching event on coral reef flats of Ishigaki Island, Ryukyu Islands. *Galaxea* **1**, 73–82 (1999) (in Japanese).
34. Diaz-Pulido, G. *et al.* Doom and boom on a resilient reef: climate change, algal overgrowth and coral recovery. *PLoS One* **4**, e5239, doi: 10.1371/journal.pone.0005239 (2009).
35. van Oppen, M. J., Lukoschek, V., Berkelmans, R., Peplow, L. M. & Jones, A. M. A population genetic assessment of coral recovery on highly disturbed reefs of the Keppel Island archipelago in the southern Great Barrier Reef. *PeerJ* **3**, e1092, doi: 10.7717/peerj.1092 (2015).
36. Davey, J. W. & Blaxter, M. L. RADSeq: next-generation population genetics. *Brief Funct Genomics* **9**, 416–423, doi: 10.1093/bfpg/elq031 (2010).
37. Combosch, D. J. & Vollmer, S. V. Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific Pocillopora corals. *Mol Phylogenet Evol* **88**, 154–162, doi: 10.1016/j.ympev.2015.03.022 (2015).
38. Wang, S., Meyer, E., McKay, J. K. & Matz, M. V. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature methods* **9**, 808–810, doi: 10.1038/nmeth.2023 (2012).
39. Fukami, H., Omori, M. & Hata, M. Phylogenetic relationships in the coral family acroporidae, reassessed by inference from mitochondrial genes. *Zoolog Sci* **17**, 689–696, doi: 10.2108/zsj.15.689 (2000).
40. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579, doi: 10.1093/bioinformatics/btq683 (2011).
41. Li, R. *et al.* The sequence and *de novo* assembly of the giant panda genome. *Nature* **463**, 311–317, doi: 10.1038/nature08696 (2010).
42. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, doi: 10.1093/bioinformatics/btp324 (2009).
43. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, doi: 10.1093/bioinformatics/btp352 (2009).
44. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303, doi: 10.1101/gr.107524.110 (2010).
45. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nature protocols* **5**, 1564–1573, doi: 10.1038/nprot.2010.116 (2010).

46. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559–575, doi: 10.1086/519795 (2007).
47. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190, doi: 10.1371/journal.pgen.0020190 (2006).
48. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904–909, doi: 10.1038/ng1847 (2006).
49. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158, doi: 10.1093/bioinformatics/btr330 (2011).
50. R: A Language and Environment for Statistical Computing (2014).
51. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* **8**, e1002967, doi: 10.1371/journal.pgen.1002967 (2012).
52. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093, doi: 10.1534/genetics.112.145037 (2012).
53. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494, doi: 10.1038/nature08365 (2009).

Acknowledgements

This study was supported in part by JSPS KAKENHI (24241071 to Noriyuki Satoh, 25660172 and 26290065 to Chuya Shinzato), and by generous support from OIST Graduate University to the Marine Genomics Unit. This project was also conducted as a part of the Coral Reef Conservation and Restoration Project by Okinawa Prefecture, Japan. We thank IDEA Consultants, Inc., for helping with the collection of *A. digitifera* specimens from the Kerama Islands. We thank three anonymous reviewers for insightful critiques that improved the manuscript. We thank Dr. Steven D. Aird (OIST) for editing the manuscript.

Author Contributions

S.M. and N.A. produced sequencing data, and C.S. and S.M. analyzed the data. C.S. and N.S. wrote the main manuscript text.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Shinzato, C. *et al.* Genome-wide SNP analysis explains coral diversity and recovery in the Ryukyu Archipelago. *Sci. Rep.* **5**, 18211; doi: 10.1038/srep18211 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>