LETTER

doi:10.1038/nature10249

Using the *Acropora digitifera* genome to understand coral responses to environmental change

Chuya Shinzato[1]*, Eiichi Shoguchi[1]*, Takeshi Kawashima[1]*, Mayuko Hamada[1]*, Kanako Hisata[1], Makiko Tanaka[1], Manabu Fujie[2], Mayuki Fujiwara[1], Ryo Koyanagi[1], Tetsuro Ikuta[1], Asao Fujiyama[3], David J. Miller[4] & Nori Satoh[1]

Despite the enormous ecological and economic importance of coral reefs, the keystone organisms in their establishment, the scleractinian corals, increasingly face a range of anthropogenic challenges including ocean acidification and seawater temperature rise[1–4]. To understand better the molecular mechanisms underlying coral biology, here we decoded the approximately 420-megabase genome of *Acropora digitifera* using next-generation sequencing technology. This genome contains approximately 23,700 gene models. Molecular phylogenetics indicate that the coral and the sea anemone *Nematostella vectensis* diverged approximately 500 million years ago, considerably earlier than the time over which modern corals are represented in the fossil record (~240 million years ago)[5]. Despite the long evolutionary history of the endosymbiosis, no evidence was found for horizontal transfer of genes from symbiont to host. However, unlike several other corals, *Acropora* seems to lack an enzyme essential for cysteine biosynthesis, implying dependency of this coral on its symbionts for this amino acid. Corals inhabit environments where they are frequently exposed to high levels of solar radiation, and analysis of the *Acropora* genome data indicates that the coral host can independently carry out *de novo* synthesis of mycosporine-like amino acids, which are potent ultraviolet-protective compounds. In addition, the coral innate immunity repertoire is notably more complex than that of the sea anemone, indicating that some of these genes may have roles in symbiosis or coloniality. A number of genes with putative roles in calcification were identified, and several of these are restricted to corals. The coral genome provides a platform for understanding the molecular basis of symbiosis and responses to environmental changes.

Coral reefs are estimated to harbour around one third of all described marine species[6], and their productivity supports around one quarter of marine fisheries, but declines in coral abundance and wholesale loss of reef habitats are one of the most pressing environmental issues of our time. The major architects of coral reefs, the scleractinian corals, are anthozoan cnidarians that form obligate endosymbioses with photosynthetic dinoflagellates of the genus *Symbiodinium* (Fig. 1b). The symbionts confer on the coral holobiont the ability to fix CO_2 and to deposit the massive aragonite (a form of calcium carbonate) skeletons that distinguish reef-building corals from other anthozoans such as sea anemones. The association is fragile, however, and collapses under stress. Despite the ecological and economic significance of corals, the molecular mechanisms underlying much of coral biology—including stress responses and disease—remain unclear, but it is clear that corals retain much of the complex gene repertoire of the ancestral metazoan[7]. To address the lack of molecular data for reef-building corals, we determined the whole-genome sequence of *A. digitifera* (Fig. 1a–h), a dominant species on Okinawan reefs. Not only are *Acropora* species the dominant reef-building corals of the Indo-Pacific, but they are also among the most sensitive of corals to increased seawater temperatures[8].

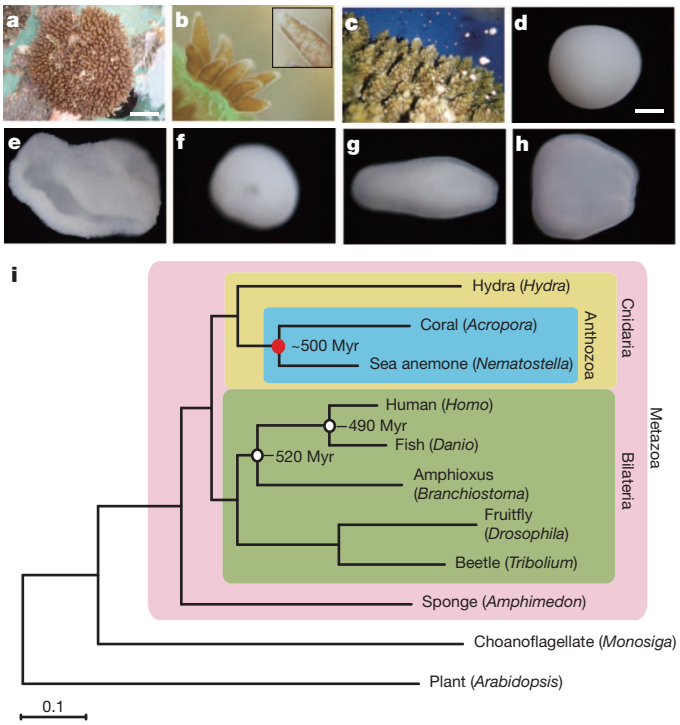

Figure 1 | The coral *Acropora digitifera* and an early occurrence of corals on Earth. a, The colony the genome of which was sequenced in the present study. This colony is maintained in aquarium culture at the Sesoko Station, University of the Ryukyus, and is thus available for further investigation of the genome. Scale bar, 10 cm. **b**, Polyps of the coral showing the presence of symbiotic dinoflagellates (*Symbiodinium* sp.) (inset, enlargement). **c**, Natural spawning of the coral. **d**–**h**, Eggs, embryos, larva and primary polyp of *A. digitifera*, from which messenger RNA was extracted for transcriptome analyses. Scale bar, 200 μm. **d**, Fertilized egg; **e**, blastula at the prawn chip stage; **f**, gastrula; **g**, planula larva; and **h**, primary polyp. **i**, Molecular phylogeny of corals. 94,200 aligned amino acid positions of proteins encoded by 422 genes were obtained from the sponge *Amphimedon queenslandica*, from the cnidarians *A. digitifera*, *Nematostella vectensis* and *Hydra magnipapillata*, and from the triploblasts *Tribolium castaneum*, *Drosophila melanogaster*, *Branchiostoma floridae*, *Danio rerio* and *Homo sapiens*. The sequences were analysed using maximum likelihood methods, with the plant *Arabidopsis thaliacia* and the choanoflagellate *Monosiga brevicollis* serving as outgroups. The scale bar represents 0.1 expected substitutions per site in the aligned regions. The topology was supported by 100% bootstrap value. Approximate divergent times of the occurrence of basal chordates and divergence of vertebrates lineages are shown. This analysis indicates a deeper divergence of *Acropora* and *Nematostella*, approximately 500 million years (Myr) ago.

[1]Marine Genomics Unit, Okinawa Institute of Science and Technology Promotion Corporation, Onna, Okinawa 904-0412, Japan. [2]DNA Sequencing Center Section, Okinawa Institute of Science and Technology Promotion Corporation, Onna, Okinawa 904-0412, Japan. [3]National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan. [4]ARC Centre of Excellence for Coral Reef Studies and School of Pharmacy and Molecular Sciences, James Cook University, Townsville, Queensland 4811, Australia.
*These authors contributed equally to this work.

On the basis of flow cytometry, the *A. digitifera* genome is approximately 420 Mbp (Supplementary Figs 1 and 2) and is therefore similar in size to that of the sea anemone *Nematostella*[9]. Sperm from a single colony served as the source of DNA for sequencing using a combination of Roche 454 GS-FLX[10] and Illumina Genome Analyser IIx (GAIIx)[11] methods. The genome was sequenced to approximately 151-fold coverage (Supplementary Table 1), enabling the generation of an assembly comprising a total of 419 Mbp (Supplementary Tables 2–5; contig N50 = 10.7 kbp and scaffold N50 = 191.5 kbp; Supplementary Fig. 3). The genome is approximately 39% G+C (Supplementary Fig. 4), and contains 23,668 predicted protein-coding loci (Supplementary Table 6). Transposable elements occupy approximately 12.9% of the genome (Supplementary Table 7). The coral gene set is comparable in size and composition with those of *Nematostella vectensis*[9] and *Hydra magnipapillata*[12] (Supplementary Tables 6, 8 and 9). The genome browser is accessible at http://marinegenomics.oist.jp/acropora_digitifera (Supplementary Fig. 5). Approximately 93% of the *A. digitifera* genes have matches in other metazoans (Supplementary Fig. 6a), and of these, 11% have clear homology only among expressed sequence tag (EST) data from corals[13] (Supplementary Fig. 6b), suggesting the presence of a considerable number of coral-specific genes.

Corals are morphologically very similar to sea anemones, but their evolutionary origins are obscure. Reef-building Scleractinia first appeared in the fossil record in the mid-Triassic (approximately 240 million years ago)[5], but were already highly diversified, suggesting much earlier origins. The availability of fully sequenced genomes for three cnidarians—*Acropora* (the present study), *Nematostella*[9] and *Hydra*[12]— allowed the estimation of the depth of the divergence between corals and other metazoans. Molecular phylogenetic analyses based on an alignment of 94,200 amino acid positions suggest a divergence time of *Acropora* and *Nematostella* between 520 to ~490 million years ago (the late Cambrian or early Ordovician) (Fig. 1i). The implied earlier origin of Scleractinia indicates that corals have persisted through previous periods of major environmental change, including the mass extinction event at the Permian/Triassic boundary, when global CO_2 and temperature were much higher than at present. However, whereas the Scleractinia as a lineage has persisted on evolutionary time scales, whether modern coral reefs can adapt to rapid environmental change on ecological time scales is a very different question.

The obligate endosymbiosis of corals dates at least from the mid-Triassic, and the longevity of this association might therefore be expected to have resulted in changes within the coral genome. We were unable to find any *Symbiodinium* DNA sequences in the coral genome, hence there is as yet no evidence for horizontal gene transfer from symbiont to host (Supplementary Fig. 6). However, comparative analyses indicated that, in the case of *Acropora*, the coral host might be metabolically dependent on the symbiont. Using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database[14], the metabolic repertoire of *Acropora* was compared to that of its non-symbiotic relative, the sea anemone *Nematostella* (Supplementary Table 10), leading to the identification of an apparent metabolic deficiency in *Acropora*. The biosynthesis of cysteine from homocysteine and/or serine requires the activities of two enzymes, cystathionine β-synthase (Cbs) and cystathionase (cystathionine γ-lyase; Cth) (Table 1). Whereas we were able to identify genes encoding the latter in both *A. digitifera* and *Nematostella*, the former could not be identified in *Acropora* despite a clear match being present in *Nematostella* (Supplementary Fig. 7). Although extensive transcriptomic data are available for various *Acropora* spp[13], we could find no evidence for a *Cbs* transcript in any of these. Moreover, whereas a polymerase chain reaction (PCR) strategy confirmed the presence of *Cbs* in some other corals, *Galaxea fascicularis*, *Favites chinenis*, *Favia lizardensis* and *Ctenactis echinata*, no amplification products could be obtained for two different *Acropora* species (Table 1 and Supplementary Fig. 8). Although the analyses presented here do not rigorously exclude the presence of Cbs activity in *Acropora*, they raise the intriguing possibility of a metabolic basis for the obligate nature of symbiosis in *Acropora*; differences in dependency could potentially explain not only the phenomenon of symbiont selectivity, but also the high sensitivity of *Acropora* to environmental challenges.

Table 1 | The presence or absence of a gene encoding Cbs for L-cysteine biosynthesis

	L-Homocysteine + L-Serine	*Cbs* →	L-Cystathionine	*Cth* →	L-Cysteine
Hydra magnipapillata		Yes*		Yes	
Nematostella vectensis		Yes*		Yes	
Complexa					
Acropora digitifera		No†		Yes	
Acropora tenuis		No‡		ND	
Acropora millepora		No§		Yes	
Galaxea fascicularis		Yes‡		ND	
Robusta					
Montastraea faveolata		Yes§		Yes	
Favia lizardensis		Yes‡		ND	
Favites chinensis		Yes‡		ND	
Ctenactis echinata		Yes‡		ND	

ND, not determined.
* Supported by sequenced genome and EST analyses.
† Supported by sequenced genome, EST and PCR amplification of genomic DNA.
‡ Supported by PCR amplification of genomic DNA.
§ Supported by EST analyses.

Reef-building corals typically inhabit shallow and relatively clear tropical waters and are therefore constantly exposed to high levels of ultraviolet irradiation. As corals are particularly susceptible to bleaching when exposed to both raised temperatures and high solar radiation[2,4], one intriguing question is how corals protect themselves against ultraviolet damage. Photo-protective compounds, such as the mycosporine amino acids (MAAs), have been isolated from corals[15,16] but, because similar compounds have been identified in algae, the sources of these compounds were unknown. Recently a short (four-step) pathway encoded by a gene cluster (DHQS-like, *O-MT*, ATP-grasp and NRPS-like) (Fig. 2 and Supplementary Figs 9–12) has been demonstrated to be both necessary and sufficient in the cyanobacterium *Anabaena variabilis* to convert pentose-phosphate metabolites to shinorine, a photo-protective MAA[17]. Scanning the available whole-genome data allowed us to identify clear homologues of all four members of the cyanobacterial shinorine gene cluster in both *A. digitifera* and *N. vectensis* (Fig. 2), indicating that both the coral and the sea anemone have the ability to carry out *de novo* synthesis of ultraviolet-protective compounds. Hence, MAA synthesis in corals and other cnidarians is not symbiont dependent.

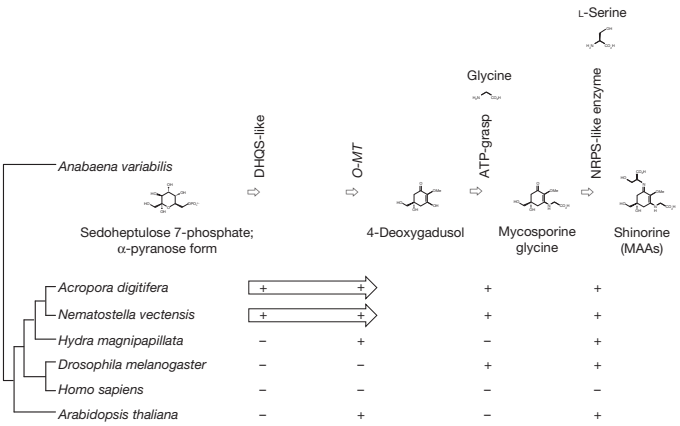

Figure 2 | The genes required for the biosynthesis of shinorine are present in anthozoan cnidarians. Top, the organization of the gene cluster involved in the biosynthetic pathway of the photo-protective molecule shinorine, a mycosporine-like amino acid, in the cyanobacterium *Anabaena variabilis*. Bottom, the presence of corresponding genes in various organisms is indicated (+). The *Acropora* and *Nematostella* genomes contain homologues of each of the four genes, in which DHQS-like and *O-MT* are fused with each other.

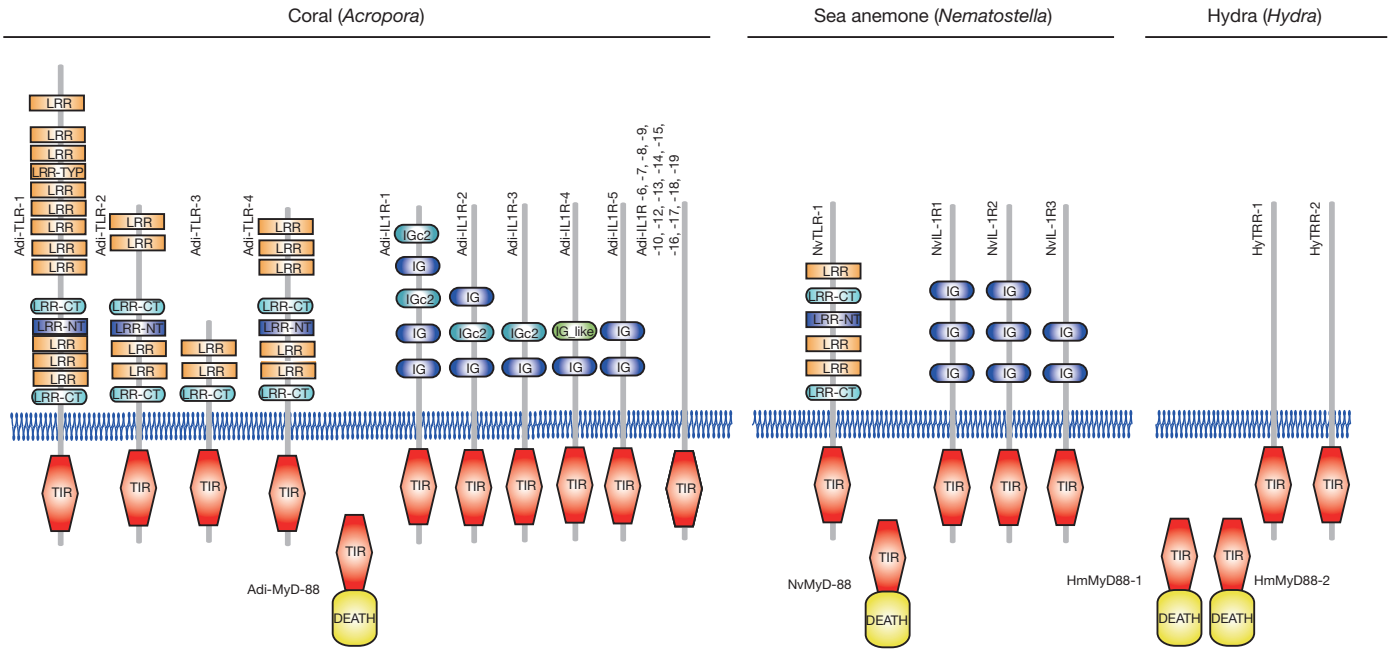

Figure 3 | Repertoires of TIR-domain-containing proteins of three cnidarians. The schematic representation of the domain structures of TIR-domain-containing proteins identified in *A. digitifera*, alongside the corresponding complements from *Nematostella vectensis* and *Hydra magnipapillata*. The repertoire of Toll/TLR, IL-1R-like and TIR-only proteins is significantly more complex in the case of *A. digitifera* than in *N. vectensis* or *H. magnipapillata*. TIR, TIR domain. DEATH, DEATH domain. IG and IGc2, Ig domain. LRR, LRY-TRY, LRR-CT and LRR-NT, leucine-rich repeats.

Surveys of *Acropora* for genes associated with innate immunity[18], apoptosis[19] and autophagy[19] indicate not only the complexity of these systems in *Acropora* (Supplementary Figs 13–23), but also that the coral innate immune repertoire is more sophisticated than that of *Nematostella*. For example, whereas a single canonical Toll/TLR protein is present in *N. vectensis*[18], the *Acropora* genome encodes at least four such molecules, as well as five IL-1R-related proteins and a number of TIR-only proteins (Fig. 3). Likewise, the *Acropora* repertoire of NACHT/NB-ARC domains, which are characteristic of primary intracellular pattern receptors[20], is again highly complex: an order of magnitude more NACHT/NB-ARC domains are present in coral than in other animals (Supplementary Table 11), and some of these cnidarian proteins have novel domain structures (Supplementary Fig. 23b). In terms of the apparent expansion and divergence of NACHT-encoding genes, the coral resembles amphioxus[21], the sea urchin[22] and angiosperms[23]. The greater complexity of the coral innate immunity network may in part reflect adaptations associated with the symbiotic state and coloniality.

The coral repertoire of genes with predicted roles in skeleton deposition is of particular interest given the likely impact of ocean acidification resulting from rising atmospheric CO_2 on coral calcification. Surveys of the *Acropora* genome for specific groups of proteins associated with calcification, including the eukaryotic-type carbonic anhydrases[24] are given in Supplementary Table 12. In general, the soluble fraction of the organic matrix in scleractinian corals is very rich in acidic amino acids, and has a particularly high aspartic acid composition[25]. A number of candidate organic-matrix proteins were identified in *Acropora* (Supplementary Fig. 24). For several of these, orthologues could be identified in *A. millepora* and/or *A. palmata* but only one of these (Adi-SAP6) was found in other coral species (Supplementary Table 13). Galaxins, first purified from the coral *Galaxea fascicularis*, are unique to corals and are the only coral skeletal matrix protein for which the complete primary structure has been determined[26]. However, galaxin possesses neither acidic regions (the fraction of Asp+Asn in the galaxin is 9.7%) nor obvious Ca^{2+}-binding domains[26]. Four genes encoding galaxin-related proteins were identified in the *A. digitifera* genome (Supplementary Fig. 25), including two likely *A. digitifera* homologues of Gfa-galaxin.

Here we decoded the 420-Mbp genome of the reef-building coral *Acropora digitifera*, with the aim of providing a platform for understanding the molecular basis of symbiosis and responses to environmental change. Some of the main findings are: (1) a relatively deep divergence of the lineage leading to the reef-building corals; (2) although we could find no evidence for horizontal gene transfer from symbiont to coral despite the long evolutionary history of the association, *Acropora* may have lost a gene essential for cysteine biosynthesis and thus be metabolically dependent on its symbionts; (3) the coral host has the ability to independently carry out *de novo* synthesis of the MAA family of photoprotective compounds; (4) the innate immune repertoire of coral is highly complex in comparison with the non-symbiotic and solitary sea anemone *Nematostella*; and (5) a number of coral-specific gene families are likely to have evolved in the context of calcification. These data also provide a basis for systems biology approaches to understanding the establishment, function and collapse of coral symbioses. If and when a whole-genome sequence becomes available for the dinoflagellate symbiont of corals *Symbiodinium* sp. (zooxanthellae), these resources will together provide additional perspectives on the symbiosis and a powerful resource for understanding the response of the holobiont to environmental stresses such as raised seawater temperatures or ocean acidification.

METHODS SUMMARY

Sperm DNA obtained from a single colony of the coral *Acropora digitifera* was used for genome sequencing by Roche 454 GS-FLX[10] and Illumina Genome Analyser IIx (GAIIx)[11]. The 454 shotgun and paired-end reads were assembled *de novo* by GS *De novo* Assembler version 2.3 (Newbler, Roche)[10], and subsequent scaffolding was performed by SOPRA[27] and SSPACE[28] using the Illumina mate-pair information. Transcriptome analysis was also performed. A set of gene model predictions (the *A. digitifera* Gene Model v. 1) was generated mainly by AUGUSTUS[29], and a genome browser has been established using the Generic Genome Browser (GBrowser) 2.17. The annotation and identification of *Acropora* genes were performed by three approaches, individual methods or combinations of the methods:

reciprocal BLAST analyses, screening the gene models against the Pfam database[30] and phylogenetic analyses.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 January; accepted 3 June 2011.
Published online 24 July 2011.

1. Hughes, T. P. *et al.* Climate change, human impacts, and the resilience of coral reefs. *Science* **301,** 929–933 (2003).
2. Hoegh-Guldberg, O. *et al.* Coral reefs under rapid climate change and ocean acidification. *Science* **318,** 1737–1742 (2007).
3. Carpenter, K. E. *et al.* One-third of reef-building corals face elevated extinction risk from climate change and local impacts. *Science* **321,** 560–563 (2008).
4. Weis, V. M. Cellular mechanisms of cnidarian bleaching: stress causes the collapse of symbiosis. *J. Exp. Biol.* **211,** 3059–3066 (2008).
5. Stanley, G. D. Jr & Fautin, D. G. Paleontology and evolution. The origins of modern corals. *Science* **291,** 1913–1914 (2001).
6. Wilkinson, C. *Status of Coral Reefs of the World* (Australian Institute of Marine Studies, 2004).
7. Miller, D. J., Ball, E. E. & Technau, U. Cnidarians and ancestral genetic complexity in the animal kingdom. *Trends Genet.* **21,** 536–539 (2005).
8. Loya, Y. *et al.* Coral bleaching: the winners and the losers. *Ecol. Lett.* **4,** 122–131 (2001).
9. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317,** 86–94 (2007).
10. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).
11. Bentley, D. R. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **16,** 545–552 (2006).
12. Chapman, J. A. *et al.* The dynamic genome of *Hydra*. *Nature* **464,** 592–596 (2010).
13. Hemmrich, G. & Bosch, T. C. Compagen, a comparative genomics platform for early branching metazoan animals, reveals early origins of genes regulating stem cell differentiation. *Bioessays* **30,** 1010–1018 (2008).
14. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36,** D480–D484 (2008).
15. Shick, J. M. & Dunlap, W. C. Mycosporine-like amino acids and related Gadusols: biosynthesis, acumulation, and UV-protective functions in aquatic organisms. *Annu. Rev. Physiol.* **64,** 223–262 (2002).
16. Rastogi, R. P. *et al.* Photoprotective compounds from marine organisms. *J. Ind. Microbiol. Biotechnol.* **37,** 537–558 (2010).
17. Balskus, E. P. & Walsh, C. T. The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. *Science* **329,** 1653–1656 (2010).
18. Miller, D. J. *et al.* The innate immune repertoire in cnidaria—ancestral complexity and stochastic gene loss. *Genome Biol.* **8,** R59 (2007).
19. Dunn, S. R., Schnitzler, C. E. & Weis, V. M. Apoptosis and autophagy as mechanisms of dinoflagellate symbiont release during cnidarian bleaching: every which way you lose. *Proc. R. Soc. Lond. B* **274,** 3079–3085 (2007).
20. Meylan, E., Tschopp, J. & Karin, M. Intracellular pattern recognition receptors in the host response. *Nature* **442,** 39–44 (2006).
21. Huang, S. *et al.* Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res.* **18,** 1112–1126 (2008).
22. Hibino, T. *et al.* The immune gene repertoire encoded in the purple sea urchin genome. *Dev. Biol.* **300,** 349–365 (2006).
23. Leister, D. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. *Trends Genet.* **20,** 116–122 (2004).
24. Jackson, D. J., Macis, L., Reitner, J., Degnan, B. M. & Wörheide, G. Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis. *Science* **316,** 1893–1895 (2007).
25. Sarashina, I. & Endo, K. Skeletal matrix proteins of invertebrate animals: comparative analysis of their amino acid sequences. *Paleontological Res.* **10,** 311–336 (2006).
26. Fukuda, I. *et al.* Molecular cloning of a cDNA encoding a soluble protein in the coral exoskeleton. *Biochem. Biophys. Res. Commun.* **304,** 11–17 (2003).
27. Dayarian, A., Michael, T. P. & Sengupta, A. M. SOPRA: scaffolding algorithm for paired reads via statistical optimization. *BMC Bioinformatics* **11,** 345 (2010).
28. Boetzer, M. *et al.* Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27,** 578–579 (2011).
29. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24,** 637–644 (2008).
30. Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **38,** D211–D222 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge A. Iguchi, K. Sakai and all other staff members of the Sesoko Station, Tropical Biosphere Research Center, University of the Ryukyus for their help in collection and maintenance of the coral, M. Hidaka and V. Weis for suggestions, Y. Yokoyama, T. Usami and other staff members of our Unit for sequencing, and S. Brenner, R. Baughman and T. Ichikawa for their support. D. Rokhsar and J. Chapman are acknowledged for suggestions on sequence assembly and gene prediction. The super-computing was supported by the IT Section of OIST and the Human Genome Center, University of Tokyo. This work was supported in part by Grants-in-Aids from MEXT and JSPS, Japan.

Author Contributions Project design and coordination: N.S., C.S., E.S., T.K., M.H. and D.J.M. Genome and EST sequencing and assembly: C.S., T.K., E.S., K.H., M. Fujie, M. Fujiwara, M.T., M.H., A.F. and R.K. Annotation and analysis: E.S., C.S., M.H., T.K., K.H., M.T., R.K. and T.I. Writing: N.S., C.S., E.S., T.K., M.H. and D.J.M.

Author Information The genome assembly has been deposited with the DNA Databank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL) and GenBank under project accession BACK01000001–BACK01053640 (contigs) and DF093604–DF097774 (scaffolds). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.S. (norisky@oist.jp).

METHODS

Biological specimen. Under permits from the Aquaculture Agency of Okinawa Prefecture (the number 20–27), part of an *A. digitifera* colony was collected and has subsequently been maintained in an aquarium at the Sesoko Station, Tropical Biosphere Research Center, University of the Ryukyus.

The number of chromosomes, diploidy and genome size of *Acropora digitifera*. The number of chromosomes was determined by their preparation from nuclei of embryonic cells. The diploidy of the genome was examined by fluorescent *in situ* hybridization (FISH) of BAC clones[31], which were constructed in pKS146 (ref. 32). The genome size was estimated by flow cytometry[33] using sperm nuclei from the same colony that was used to sequence the genome.

Genome sequencing and assembly. The sperm was obtained from the single colony and sperm DNA was used for genome sequencing and BAC library construction. Genome sequence data were obtained using single read, paired-end and mate-pair protocols on the Roche 454 GS-FLX[10] and Illumina GAIIx[11] instruments. The genomic DNA was fragmented, libraries prepared and sequencing conducted according to the manufacturer's protocols. The 454 shotgun and paired-end reads were assembled *de novo* by GS *De novo* Assembler version 2.3 (Newbler, Roche)[10] in heterozygotic mode with adjusted algorisms to reflect an increase in the expected variability in sequence identity. Possible PCR duplicates in Illumina mate-pair reads were removed by MarkDuplicates in Picard tools (http://picard.sourceforge.net), and then subsequent scaffolding of the 29,765 Newbler output was performed by SOPRA[27] and SSPACE[28] using the Illumina mate-pair information. Gaps inside the scaffolds were closed with Illumina paired-end data using GapCloser[34]. To overcome potential assembly errors arising from tandem repeats, sequences that were aligned to another sequence over 50% of the length by BLASTN (1×10^{-50}) were removed from the assembly[35].

Transcriptome analyses. RNA was isolated from eggs, gastrulae, planulae, polyps and adults. Total RNA was extracted following the manufacturer's instructions (Invitrogen) and purified using DNase and an RNeasy micro kit (QIAGEN). Transcriptome libraries for 454 GS-FLX were prepared[36] and sequenced as per manufacturer's instructions. In addition, Illumina 50-bp paired-end RNA-seq sequencing was performed. All high-quality sequences (quality value $\geq$15) were assembled by a Velvet/Oases assembler[37] with hash length 27.

Gene prediction. A set of gene model predictions (the *A. digitifera* Gene Model v. 1) was generated using AUGUSTUS[29]. AUGUSTUS 2.0.4 was trained on the 877 EST assemblies recommended by PASA[38] for this purpose. The gene models were created by running AUGUSTUS on a repeat-masked genome produced by RepeatMasker[39], and improved by PASA[38]. A genome browser has been established using the assembled genome sequences using the Generic Genome Browser (GBrowser) 2.17 (ref. 40).

Identification of *Acropora* genes involved in the response to environmental change. Three approaches, individual methods or combinations of the methods, were used to annotate the protein-coding genes in the *A. digitifera* genome. A primary approach to the identification of putative orthologues of *A. digitifera* genes was reciprocal BLAST analysis. This was carried out on the basis of mutual best hit in BLAST analyses for human, mouse, or *Drosophila* genes against the *A. digitifera* gene models (BLASTP) or the assembly (BLASTN). A second approach used in the case of genes encoding proteins with one or more specific protein domains, was to screen the merged models against the Pfam database (Pfam-A.hmm, release 24.0; http://pfam.sanger.ac.uk)[30], which contains 11,912 conserved domains using HMMER (hmmer3)[41]. In the case of complex multigene families, a third annotation method was used; sets of related sequences were subjected to phylogenetic analyses to determine more precisely orthology relationships. For these purposes, amino acid sequences were aligned using ClustalW[42] or ClustalX[42] under the default options. Gaps and ambiguous areas were excluded using Gblocks 0.91b[43] with the default parameters and then checked manually. On the basis of the alignment data sets, phylogenetic trees were constructed by neighbour joining and/or maximum likelihood. Calculations and tree construction were performed in SeaView[44]. The KEGG pathway database[14] was used to examine the metabolic repertoire of *Acropora* in comparison to that of the sea anemone *Nematostella*.

31. Shoguchi, E. *et al.* Chromosomal mapping of 170 BAC clones in the ascidian *Ciona intestinalis*. *Genome Res.* **16,** 297–303 (2006).
32. Fujiyama, A. *et al.* Construction and analysis of a human-chimpanzee comparative clone map. *Science* **295,** 131–134 (2002).
33. Davies, D. & Allen, P. in *Flow Cytometry: Principles and Applications* (ed. Macey, M. G.) (Humana, 2007).
34. Li, R. *et al.* The sequence and *de novo* assembly of the giant panda genome. *Nature* **463,** 311–317 (2010).
35. Wurm, Y. *et al.* The genome of the fire ant *Solenopsis invicta*. *Proc. Natl Acad. Sci. USA* **108,** 5679–5684 (2011).
36. Meyer, E. *et al.* Sequencing and *de novo* analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics* **10,** 219 (2009).
37. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–829 (2008).
38. Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31,** 5654–5666 (2003).
39. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-3.0. ⟨http://www.repeatmasker.org⟩ (1996–2010).
40. Stein, L. D. *et al.* The generic genome browser: a building block for a model organism system database. *Genome Res.* **12,** 1599–1610 (2002).
41. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14,** 755–763 (1998).
42. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).
43. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17,** 540–552 (2000).
44. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27,** 221–224 (2010).