

1 **Additional files1:**

2

3

4 **Supplementary Figures for**

5 **Diverse, abundant and novel viruses infecting "unculturable" but abundant**
6 **marine bacteria**

7

8 Zefeng Zhang¹, Feng Chen², Xiao Chu³, Hao Zhang³, Haiwei Luo³, Fang Qin¹, Zhiqiang Zhai¹,
9 Mingyu Yang¹, Jing Sun⁴, Yanlin Zhao¹*

10

11 ¹ Fujian Provincial Key Laboratory of Agroecological Processing and Safety Monitoring,
12 College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, Fujian, 350002,
13 China

14 ² Institute of Marine and Environmental Technology, University of Maryland Center for
15 Environmental Science, Baltimore, MD, USA

16 ³ Simon F. S. Li Marine Science Laboratory, School of Life Sciences and Partner State Key
17 Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong
18 SAR, China

19 ⁴ Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao,
20 Shandong, 266071, China

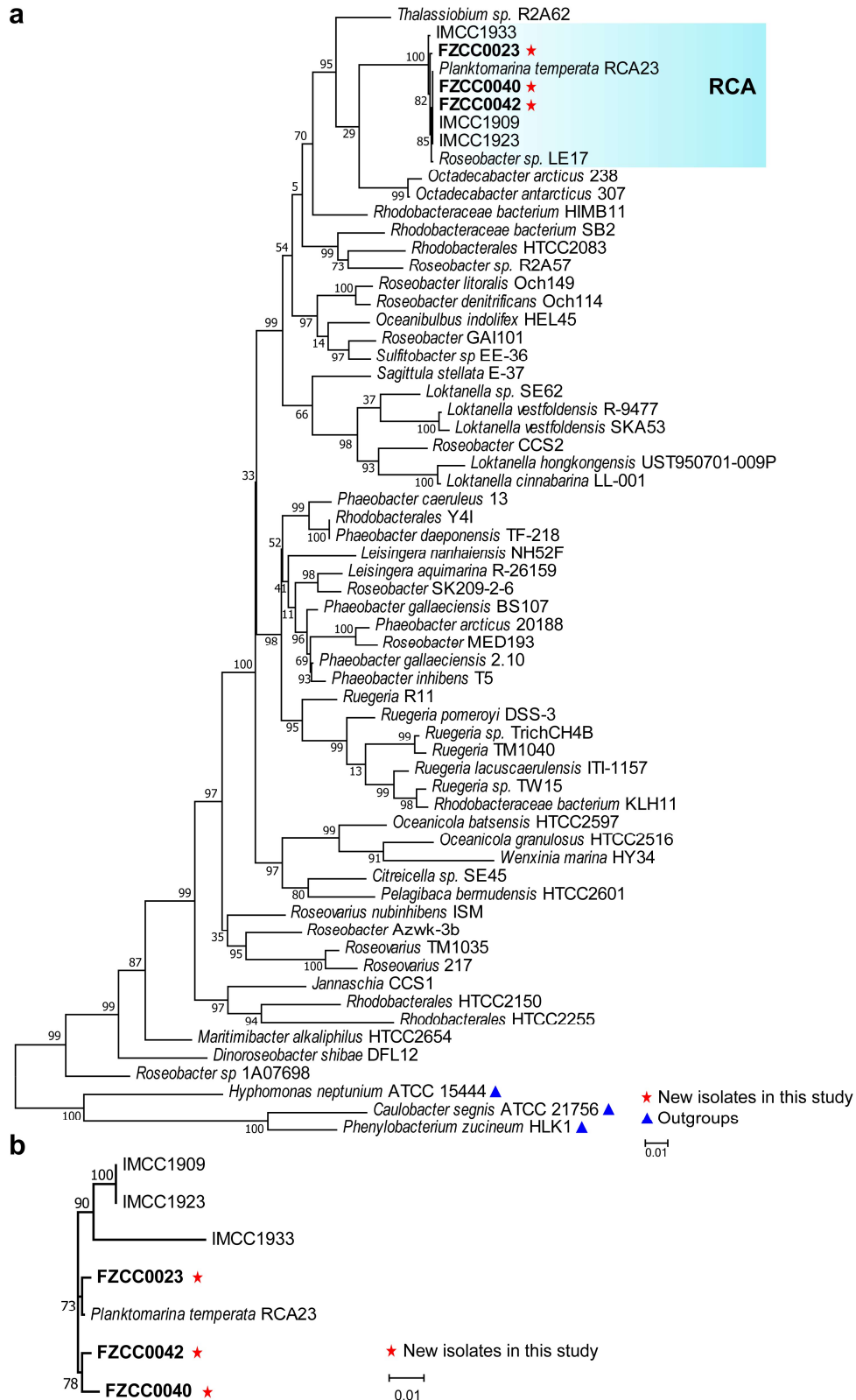
21

22 * Correspondence: yanlinzhao@fafu.edu.cn

23

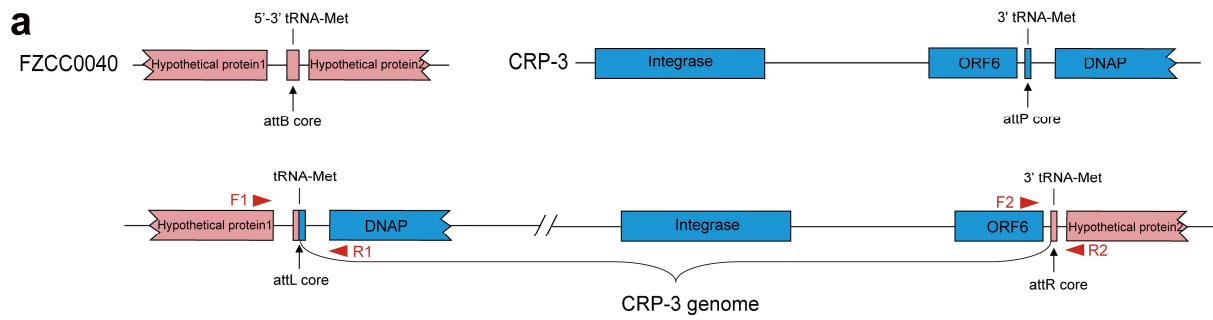
24

25



26
27
28
29
30
31

Figure S1: Phylogenetic relationship between three RCA isolates and the representatives belonging to the *Roseobacter* clade of the *Alphaproteobacteria*. **a** Phylogenetic analysis of 16S rRNA gene. **b** Phylogenetic analysis of our RCA isolates and four RCA representatives based on the ITS sequences.



b

attB: [CGCGGGATGGAGCAGCCCGGTAGCTCGTCAGGCTCATAACTGAAGGTCGTAGGTTCAAATCCTACTCCCGCAACCA]AGTGTTCG tRNA-Met(CAT)

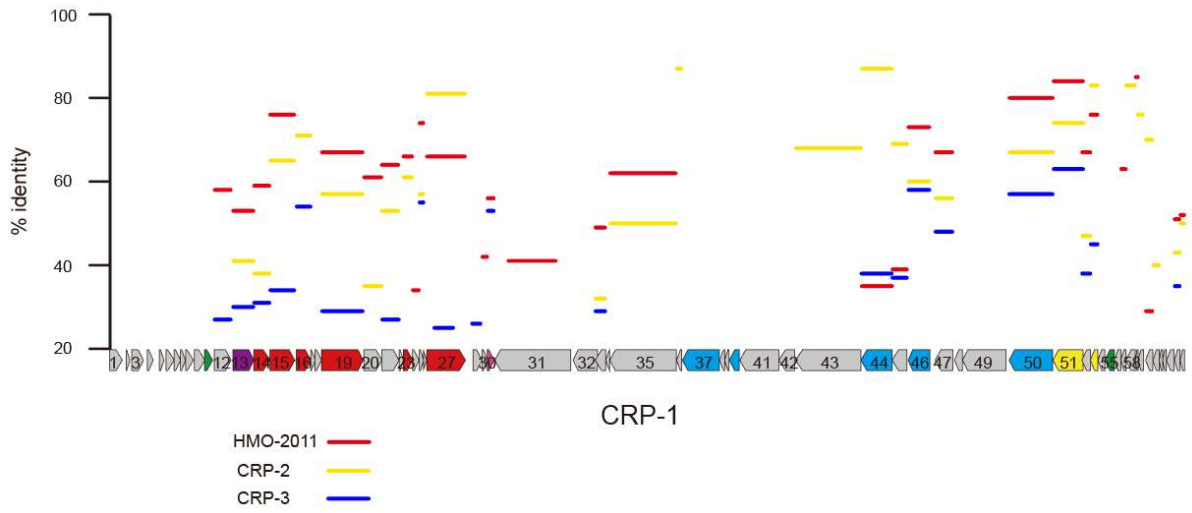
attP: TACCTTGTAAGTACCTGATATATATACTGGTTCGTATACCCTGAAGGTCGTAGGTTCAAATCCTACTCCCGCAACCAAGCCTTTGA

attL: [CGCGGGATGGAGCAGCCCGGTAGCTCGTCAGGCTCATAACTGAAGGTCGTAGGTTCAAATCCTACTCCCGCAACCA]AGCCTTTGA

attR: TACCTTGTAAGTACCTGATATATATACTGGTTCGTATACCCTGAAGGTCGTAGGTTCAAATCCTACTCCCGCAACCAAGTGTTCG

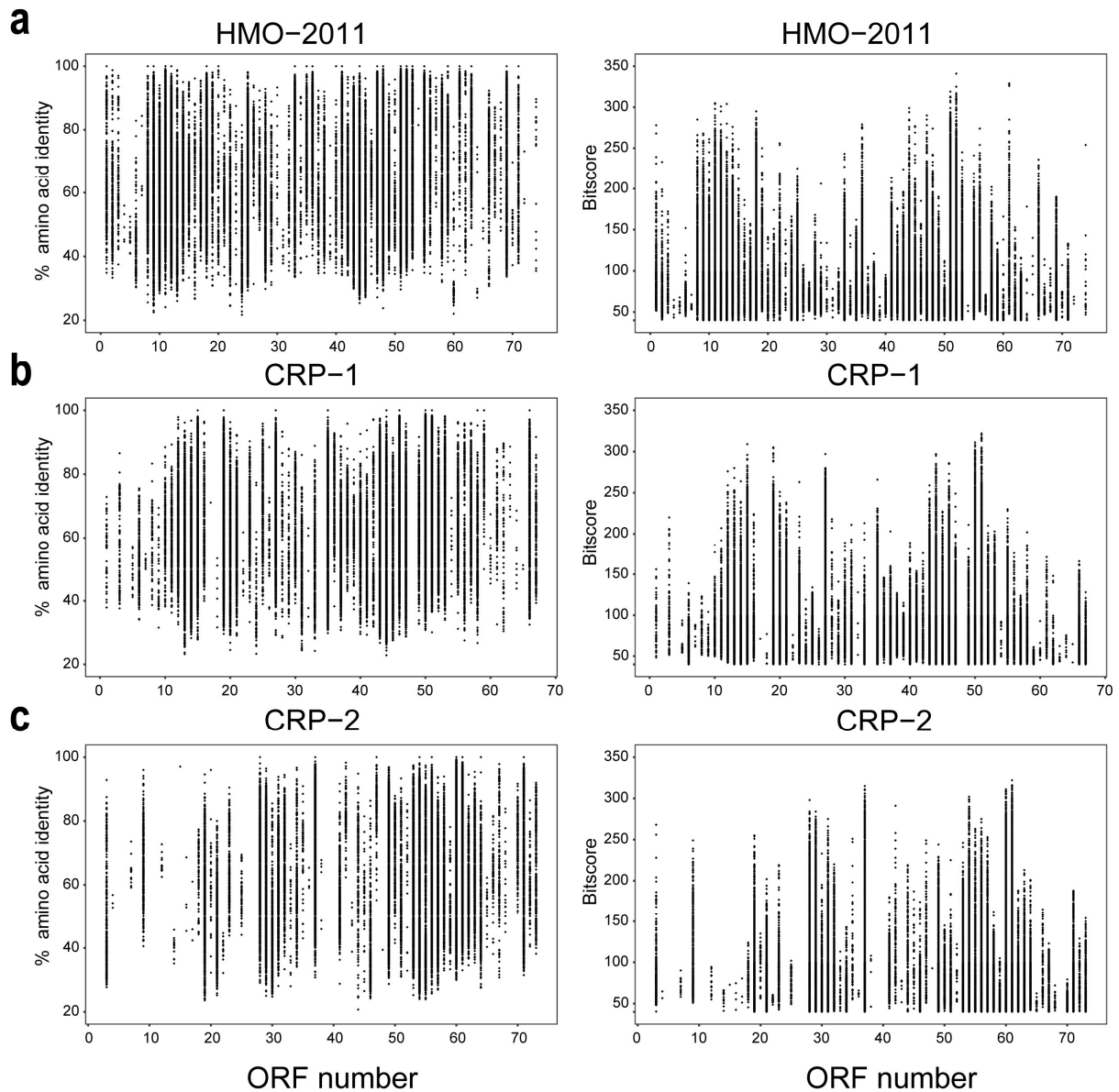
32
33
34
35
36
37
38
39

Figure S2: Integration of CRP-3 in the FZCC0040 genome. **a** CRP-3 and host FZCC0040 genes are shown in pink and blue respectively. The core sequences are indicated by the arrows. PCR primers are indicated by red triangles. **b** The DNA sequences of all integration sites. CRP-3 sequences, FZCC0040 sequences and the identical core sequences are shown in black, blue and red, respectively. The tRNA genes are boxed. Abbreviation: DNAP, DNA polymerase



40
 41
 42
 43
 44

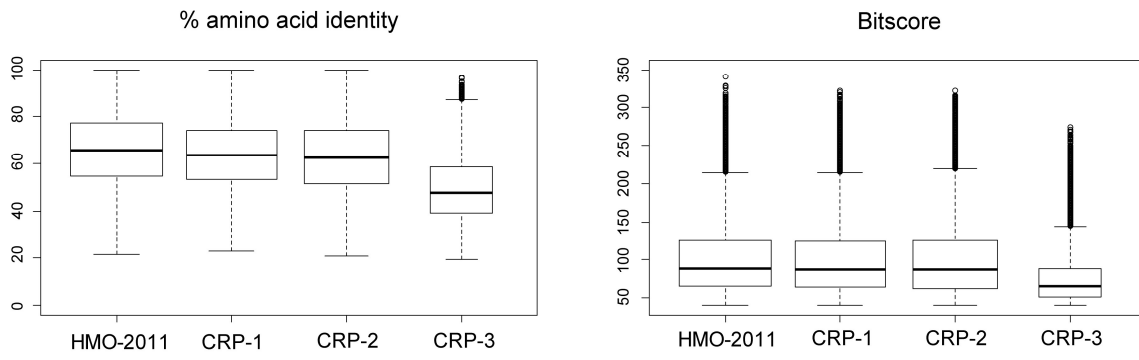
Figure S3: Percent amino acid identity of phage HMO-2011, CRP-2 and CRP-3 to the CRP-1 ORFs.



45
 46
 47
 48
 49
 50
 51
 52
 53

Figure S4: Fragment recruitment plot of % amino acid identity and bitscores of Pacific Ocean Virome reads against the ORFs of phage HMO-2011, CRP-1 and CRP-2. The total number of POV reads assigned to the *HMO-2011viruse* group is 97,684. Numbers of reads mapped to HMO-2011, CRP-1 and CRP-2 are 78,349, 74,699 and 66,560, respectively. **a**, HMO-2011. **b**, CRP-1. **c**, CRP-2

54

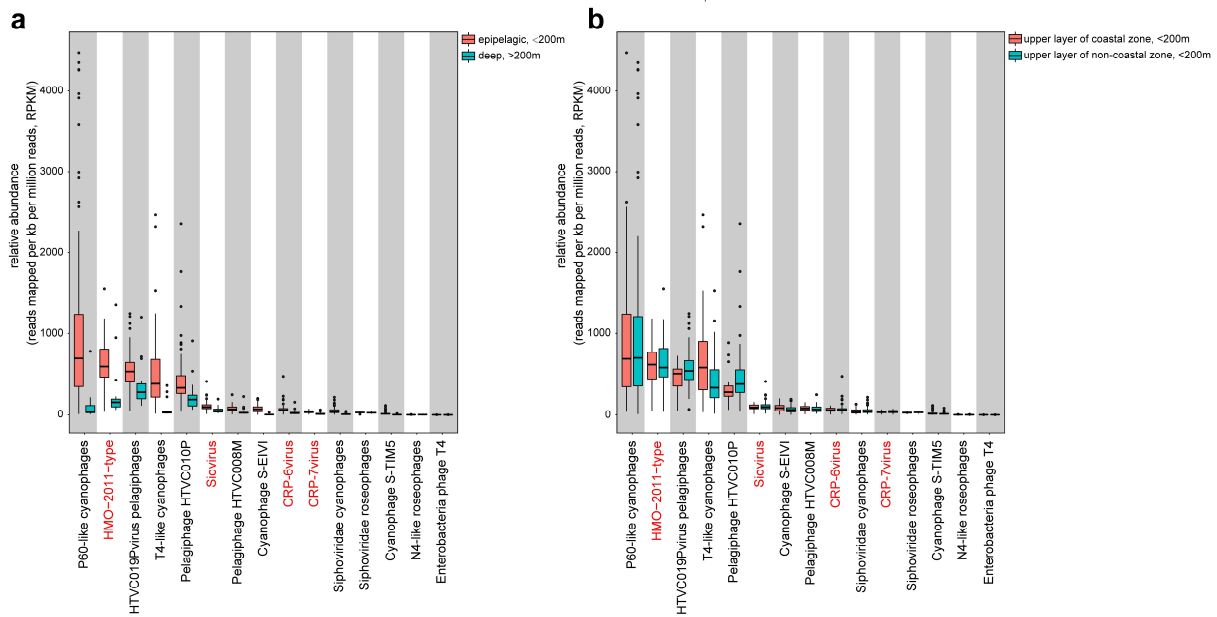


55

56

57 **Figure S5:** Boxplot distributions of % amino acid identity and bitscores of the POV reads
58 assigned to *HMO-2011virus* group against each *HMO-2011virus* genome.
59

59



61

62

63 **Figure S6 Relative abundance of major phage groups in Global Oceans Viromes (GOV).**

64 Relative abundance (y axis) of each phage group (sorted according to its median relative
 65 abundance, x axis) in each virome dataset were calculated and normalized as RPKM (the
 66 number of reads recruited per kilobase pairs in average genome size per million of reads in the
 67 virome dataset). A Relative abundance of major phage groups in upper ocean samples (water
 68 depth <200m) compared with their abundance in deep ocean samples (water depth >200m). B.
 69 Relative abundance of major phage groups in coastal water samples compared with their
 70 abundance in non-coastal water samples. The phage groups containing RCA phages were
 71 shown in red.

72