PNAS
www.pnas.org

Supplementary Information for

Historical change of El Niño properties shed light on future changes of extreme El Niño

Authors: Bin Wang, Xiao Luo, Young-Min Yang, Weiyi Sun, Mark A. Cane, Wenju Cai, Sang-Wook Yeh, and Jian Liu

Email: wangbin@hawaii.edu

This PDF file includes:

Supplementary text
Figs. S1 to S6
References for SI reference citations

Supplementary Information Text

Materials and Methods

Data

The SST data used are made from an ensemble mean of two sets of monthly mean SST data from 1871 to 2017. One is the Hadley Center Sea Ice and SST dataset version 1 (HadISST1) from 1871 to date with a resolution of 1˚×1˚ derived from the Met Office Marine Data Bank and the WMO GTS (1). HadISST1 temperatures are reconstructed using a two-stage reduced-space optimal interpolation procedure. HadISST1 provides global SST though it has sparse data input in the polar regions and the Southern Ocean, and it has non-robust trends, because changes in SST measurement practices were not considered in the post-1941era. Another is the ERSST V5 global SST monthly dataset from 1854 to date with a resolution of 2˚×2˚, which is derived mainly from ICOADS R3.0 SST and Argo floats above 5 meters (2). ERSSTv5 has improved SST spatial and temporal variability and thus provides improved absolute SST. The data prior to the 1880s in the two datasets may not be reliable due to sparse observations in the Pacific and El Niño Indian Oceans. To reduce the uncertainty in each set, we have made a "merged" monthly mean SST dataset (merged SST hereafter) from the two datasets by simply taking their arithmetic means. Figure 1, 3, 4, and 5 used the merged SST data.

The ocean reanalysis datasets used are primarily from the SODA version 2.2.4 reanalysis for 1871-2008 with a resolution of 0.5˚×0.5˚ (3,4). The ocean model is based on Parallel Ocean Program model physics and forced by the 20Crv2 winds. Virtually all available hydrographic profile data, ocean station data, and salinity time series were assimilated. During the recent years from 2009 to 2018, we used the Global Ocean Data Assimilation System (GODAS) with a 1˚×1˚ grid (5). GODAS is a real-time ocean analysis and reanalysis. The mean state of GODAS is

calibrated to SODA based on the overlapping period 1980-2008. The merged ocean temperature data from 1901-2017 were used for Figs. 2 and 4B.

The surface winds and atmospheric circulation fields were derived from two "merged" atmospheric reanalysis datasets, one from NOAA/NCEP and other from European Center for Medium-Range Weather Forecasts (ECMWF). The merged NCEP dataset was made from the NOAA-CIRES Twentieth Century Reanalysis (20CRv2c) (6) (1871-2012) and NCEP/DOE Reanalysis 2 data (7) (1979-2018). The 20CRv2c was assimilated by using NCEP GFS 2008ex model forced by SODAsi.2 SST and COBE-SST2 sea ice, and constrained mainly by observed surface pressures. The two NCEP model datasets were combined into a merged NCEP reanalysis dataset using 20CRv2c (1871-1979) and NCEP/DOE Reanalysis 2 (1979-2017). To ensure temporal consistency, the differences in monthly climatology between 20CRv2c and NCEP/DOE2 data during the overlap period 1979-2012 were used to calibrate mean state of NCEP/DOE2. The Fig. 4A used this dataset. The merged ECMWF reanalysis dataset was made from the ERA-20C reanalysis (8) (1901 to 2010), the ERA 40-year (ERA-40) (9) reanalysis (1958 to 2001) and the ERA-Interim reanalysis (10) (1979 to 2018). The ERA-20C data were assimilated by using Integrated Forecast system (IFS) version cy38r1(T159) and constrained by observed surface pressures and surface marine winds. The forcing data include sea ice concentration and SST from HadISST version 2.1.0.0 and other forcings are the same as those specified for CMIP5 (11). The three EC datasets were combined into a merged EC reanalysis dataset using ERA-20C (1901-1957), ERA-40 (1958-2001), and ERA-Interim (2002-2018). To ensure temporal consistency, the mean states of ERA-40 and ERA-Interim reanalysis datasets were calibrated by using the data during the overlap period 1979-2001 and by removing their differences in monthly climatology from ERA-20C dataset. To further reduce possible

uncertainty in the merged ERA and merged NCEP datasets, we have made an "ensemble mean" circulation reanalysis dataset by simply taking their arithmetic means. The "ensemble mean" circulation reanalysis data from 1901-2017 were used for surface wind analysis in Figs. 2 and 4A.

The precipitation data used are made from an ensemble mean of two sets of monthly mean precipitation data from 1901 to 2017. One is the Global Precipitation Climatology Center (GPCC) dataset over land with a resolution of 1˚×1˚ (12), which comprises observations over 85,000 stations; the other is the Climate Research Unit (CRU) TS v. 4.02 land precipitation with a resolution of 0.5˚×0.5˚ (13). To reduce the uncertainty in each set, we have also made a "merged" precipitation dataset (merged SST hereafter) from the two datasets by simply taking their arithmetic means. The merged precipitation data were used in Fig. S2.

Definition of El Niño years (1901-2017)

The SST anomaly averaged in the NINO 3.4 region (5˚N-5˚S, 120˚-170˚W), known as NINO 3.4 index or Oceanic Niño Index (ONI) has been commonly used to measure the intensity of ENSO events. The climatological annual variation of the monthly ONI exhibits a sharp peak around May (April-June) and a flat minimum period from September to January (Fig. S3A), implying a large seasonal change of the zonal SST gradients (the smallest in April-May and largest in October-January). Note that the ONI has large variances during October-November-December-January-February (ONDJF) with a maximum in December (Fig. S3B**)**. For this reason, we use ONDJF mean ONI to identify El Niño years. More precisely, we first computed a 3-month running mean SST anomaly for the NINO 3.4 region, and then averaged them during the 5-month from October (0) to February (1) to obtain a yearly ONI index,

which is referred to as ONDJF (0/1) ONI or ONDJF ONI for short, where 0 denotes the El Niño (La Niña) developing year.

The time series of ONDJF ONI from 1901 to 2017 derived from the merged SST dataset show a slight upward linear trend of 0.027K/decade, but not statistically significant at the 90% confidence level (p=0.83). The linearly detrended ONI changes slightly in the beginning and ending periods of the time series (Fig. S3C). In this study, both the original and the linearly detrended SST data are analyzed for comparison. The monthly anomalies are the departures from the climatological monthly means made for the 1901-2017 period.

An El Niño year is defined as when the ONDJF ONI is greater than or equals to 0.6°C. Using the linearly detrended data, 33 El Niño years are identified. The inconsistent years due to the removal of the small linear trend are all marginal events that occur in the beginning and the ending periods of the study period. For the period of 1949-2017 our definitions using the detrended dataset yield the same El Niño and La Niña years as those defined by NCEP/CPC (*http://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.ph*), except missing the 1953 and 1979 El Niño events and the 2016 La Niña event. These three are all weak events and they do not have typical phase-locking feature (i.e., mature during ONDJF). This slight difference is related to the fact that in the Climate Prediction Center (CPC) definition, the SST anomaly is the departure from a 30 year-running climatology centered around the target year and the El Niño (La Niña) year is defined when ONI in any 5 consecutive overlapping months is above (below) 0.5˚C (-0.5˚C).

Cluster analysis

The cluster analysis focuses on the temporal evolution characteristics of the onset, development and mature of ENSO events, which is depicted by the SST anomalies along the

equator averaged between 5˚S and 5˚N and from the October of the year prior to El Niño occurrence to the October of the El Niño year (the Hovmöller diagram). The K-means cluster analysis (14) was used to distinguish different space-time structure of SST anomalies of El Niño events. The data used in the cluster analysis (Fig. 1) were 3-month running mean SST anomalies based on the original SST data. The composite patterns derived from the original data (Fig. S4 left panels) and the corresponding detrended data (Fig. S4 right panels) are nearly identical.

In the K-means cluster analysis, the squared Euclidean distance was used to measure the "similarity" between each cluster member and the corresponding cluster centroid. The silhouette clustering evaluation criterion was used to evaluate the performance of cluster analysis (Fig. S5). The silhouette value for each member is a measure of the similarity between that member and other members in its own cluster, when compared to the members in other clusters. The silhouette value ranges from -1 to +1. A high silhouette value indicates that the member is well-matched to its own cluster, and poorly-matched to neighboring clusters (15). The silhouette values do not provide clear guidance as to how many clusters is optimal. We use K=4 clusters as the exemplars of different evolutionary patterns mainly based on physical meanings and its stability, acknowledging that there are some ambiguous cases when the silhouette value is negative or close to zero. For K=2, the discrimination is only in terms of onset from La Niña or warm/neutral conditions. For K=3, the SBW events emerge as a cluster, but the MCP and successive events are mixed in one cluster, and are separated only for K=4.

Ocean mixed layer heat budget equation

Heat budget analysis of the ocean mixed layer temperature tendency is used to quantify the contributions of different processes to the developments of three types of El Niño. This diagnostic equation can be derived as follows:

$$\frac{\partial T'}{\partial t} = -\left(\mathrm{V}' \cdot \nabla\overline{T} + \overline{\mathrm{V}} \cdot \nabla T' + \mathrm{V}' \cdot \nabla T'\right) + \frac{Q'_{net}}{\rho C_p H} + R$$

$$= -\left[\left(\frac{u'\partial\overline{T}}{\partial x} + \frac{\overline{u}\partial T'}{\partial x} + \frac{u'\partial T'}{\partial x}\right) + \left(\frac{v'\partial\overline{T}}{\partial y} + \frac{\overline{v}\partial T'}{\partial y} + \frac{v'\partial T'}{\partial y}\right) + \left(\frac{w'\partial\overline{T}}{\partial z} + \frac{\overline{w}\partial T'}{\partial z} + \frac{w'\partial T'}{\partial z}\right)\right] + \frac{Q'_{net}}{\rho C_p H} + R,$$

where the overbars and primes indicate climatological mean and anomalous quantities, respectively; T denotes the mixed layer temperature; $\mathrm{V} = (u, v, w)$ represents the zonal and meridional currents, and upwelling velocities, respectively; $\nabla = (\partial/\partial x,\ \partial/\partial y,\ \partial/\partial z)$ represents the three-dimensional gradient operator; $-u'\partial\overline{T}/\partial x$, $-v'\partial\overline{T}/\partial y$ and $-w'\partial\overline{T}/\partial z$ denote the advection of mean temperature by anomalous ocean currents in three directions; $-\overline{u}\partial T'/\partial x$, $-\overline{v}\partial T'/\partial y$ and $-\overline{w}\partial T'/\partial z$ represent the anomalous temperature advection by mean currents in three directions; $-u'\partial T'/\partial x$, $-v'\partial T'/\partial y$ and $-w'\partial T'/\partial z$ are the nonlinear advection terms in three directions. Q_{net} denotes the net downward heat flux at the ocean surface; ρ(=10^3 kg m^{-3}) is water density; C_p(=4000 J kg K^{-1}) is the specific heat of water; and R denotes the residual term. The mixed layer depth H is taken as a constant 50 m (16, 17) and the analysis result is not sensitive to the different mixed layer thickness, such as H= 30 m or 70 m.

CMIP 5 models' historical and future scenario runs

The historical and representative concentration pathway (RCP) 4.5 and RCP8.5 simulations from the Coupled Model Intercomparison Project phase 5 (CMIP5) were used to examine the dependence of the future change of strong basin-wide (SBW) El Niño events on the change of the mean-state zonal SST gradient in the central Pacific. We selected eight models that can capture the observed horizontal patterns and temporal evolutions of the three distinct El Niño events in their historical runs. Four models (CNRM-CM5, CCSM4, CanESM3 and GFDL-CM3) show a relative warming in the eastern Pacific than other equatorial regions, which resembles an "El Niño -like" mean-state SST change, namely the decreased mean-state zonal SST gradient. The other four models (GFDL-ESM2M, ACCESS1-3, MPI-ESM-MR, MRI-CGCM3) show a

"La Nina-like" mean-state SST change patterns with a relative warming in the western Pacific, namely the increased mean-state zonal SST gradient. To facilitate comparison, the periods of integration were 95 yrs for both the historical run (1911-2005) and the RCP4.5 and RCP8.5 runs (2006-2100). We have conducted the same cluster analysis on the historical and RCP4.5 and RCP8.5 from the eight models to identify the three El Niño events. Figure S6 shows composite evolutions of the equatorial Pacific SST anomalies in three types of El Niño in the historical run and RCP 4.5 run, respectively. The simulated spatial-temporal structures for the three types of El Niño are similar to those in observations (Fig. 1).

Figures

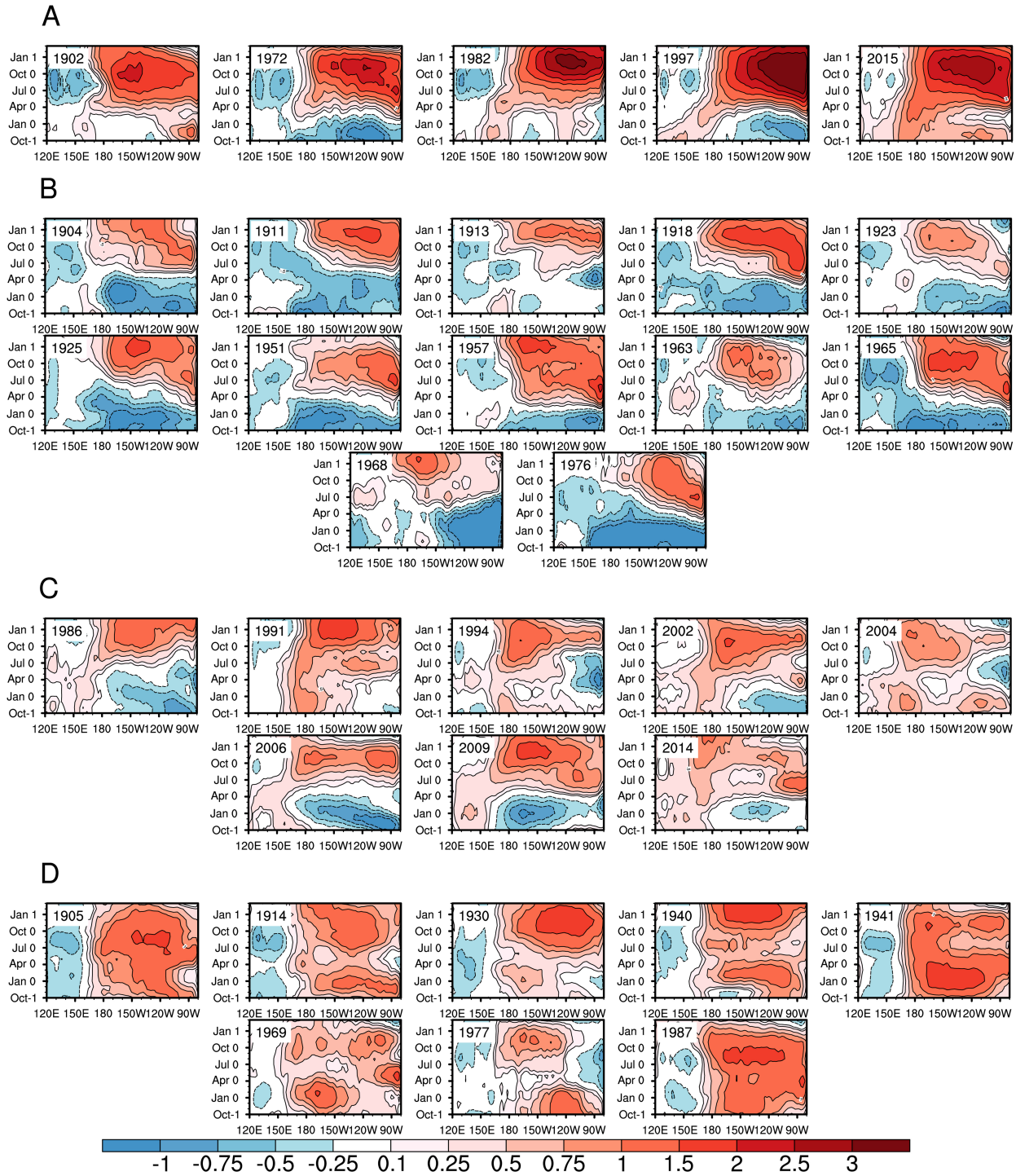

Figure S1 The equatorial SSTA evolution patterns for each individual El Niño event within each of the 4 clusters: (A) 5 strong BW, (B) 12 moderate EP, (C) 8 moderate CP, and (D) 8 Successive El Niño. The linearly detrended SSTA were used. The merged HadISST and ERSST5 data from 1901 to 2017 were used after removing small linear trends. Total of 33 El Niño events were used for cluster analysis.

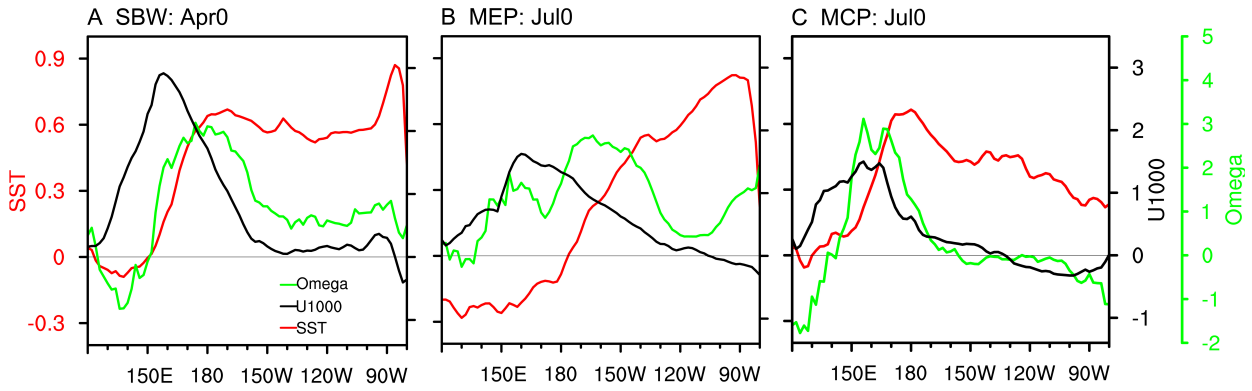

Fig. S2 Comparison of equatorial zonal structure of the three types of El Nino at their respective onset time: (A) April(0) for Strong El Nino, (B) July(0) for MEP El Nino, and (C) July(0) for MCP El Nino. The Red, black, and green lines denote SSTA (in units of ^{o}C), 1000 hPa westerly anomaly (in units of m/s), and 500 hPa vertical motion (-100hPa/s) along the equator (averaged between 5^{o}S and 5^{o}N), respectively. The figure shows the phase relationship among SSTA, zonal wind anomaly and convective anomaly.

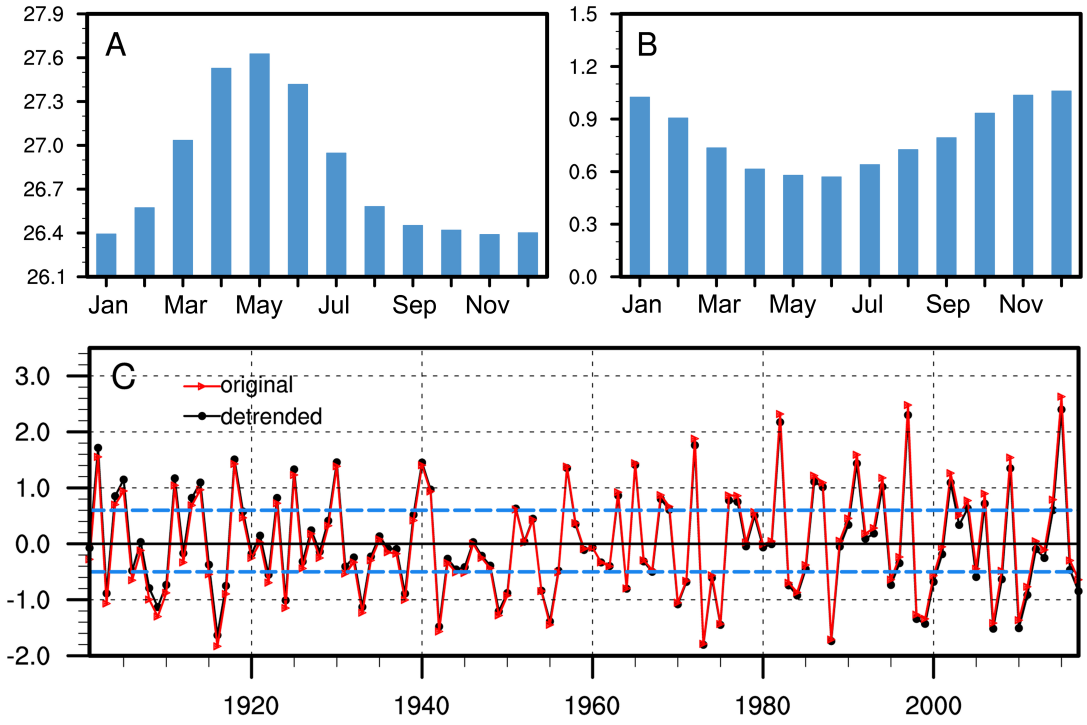

Figure S3 Climatology (1901-2017) of (A) monthly mean SST and (B) standard deviation averaged over the NINO 3.4 region (5°S-5°N, 120°W-170°W). (C) Time series of ONDJF Oceanic Niño index (ONI) (1901-2017) derived from 3-month running mean merged HadISST and ERSST5 data (red). There is an insignificant upward linear trend of 0.027k/decade (p=0.83). Black dotted line denotes the ONI after removing the linear trend. ONDJF means the five-month average from October of the current year (0) to February of the following year (1). The straight dashed blue lines of 0.6 and -0.5 indicate the criteria for identifying an El Niño and La Niña year, respectively.

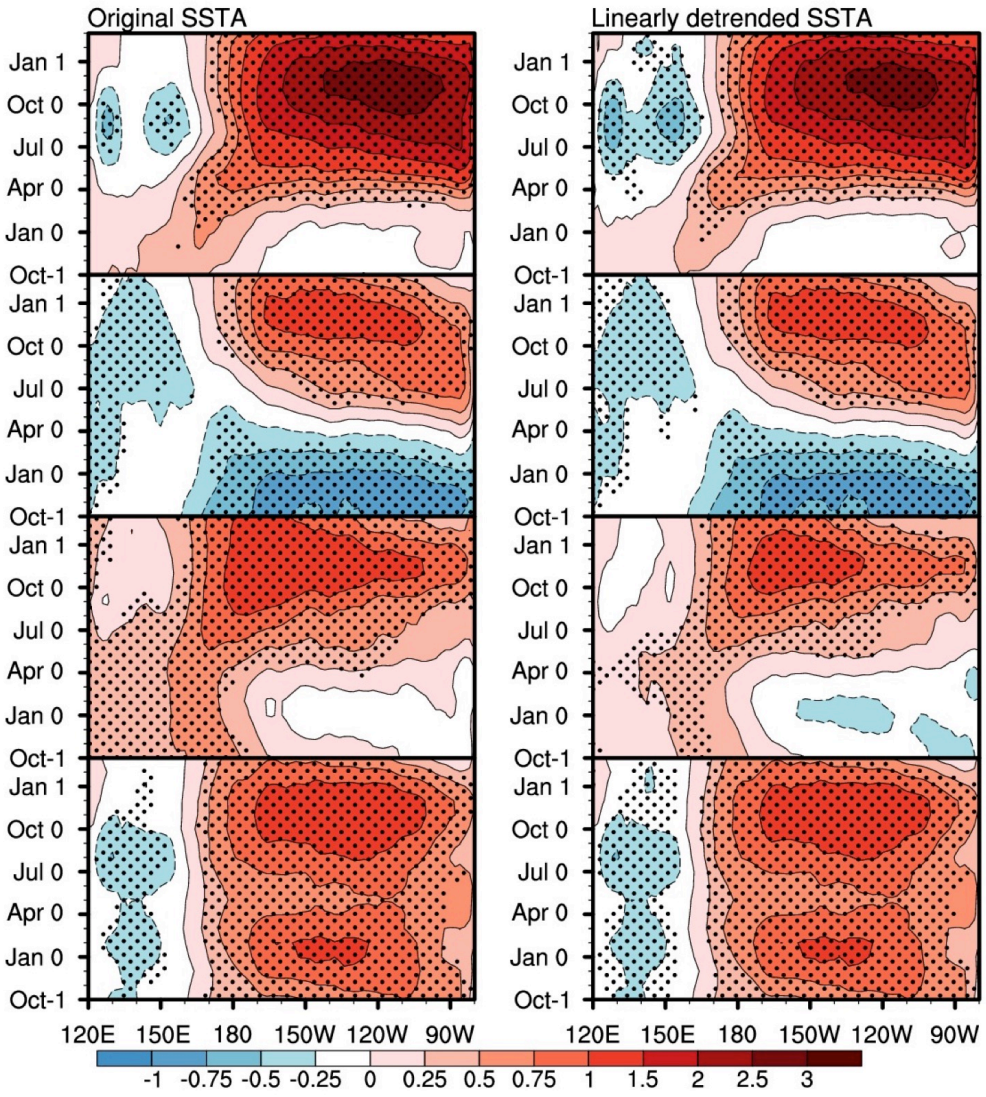

Figure S4 Composite maps of SSTA for each cluster derived from the original data (left panel) and the corresponding patterns from the detrended data (right panels) during 1901-2017 in units of °C.

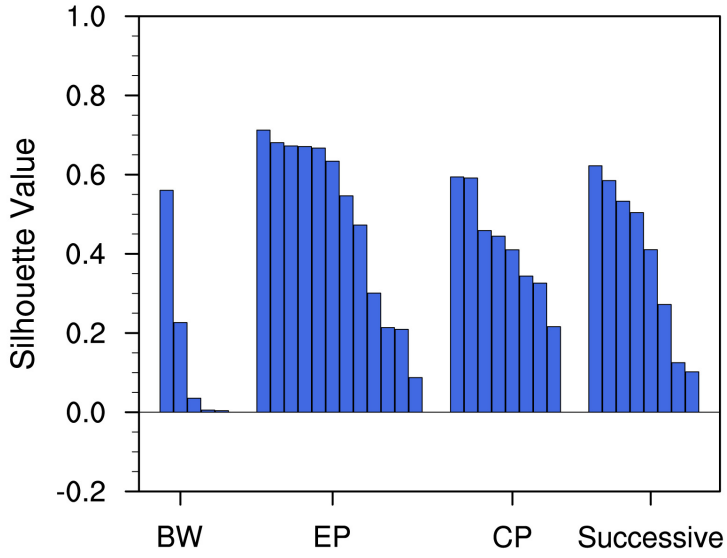

Figure S5 Silhouette values for each El Niño event within each of the 4 clusters for 1901-2017 period. The silhouette value, ranging from -1.0 to +1.0, is a measure of how similar a member is to other members in its own cluster when compared to the members in other clusters.

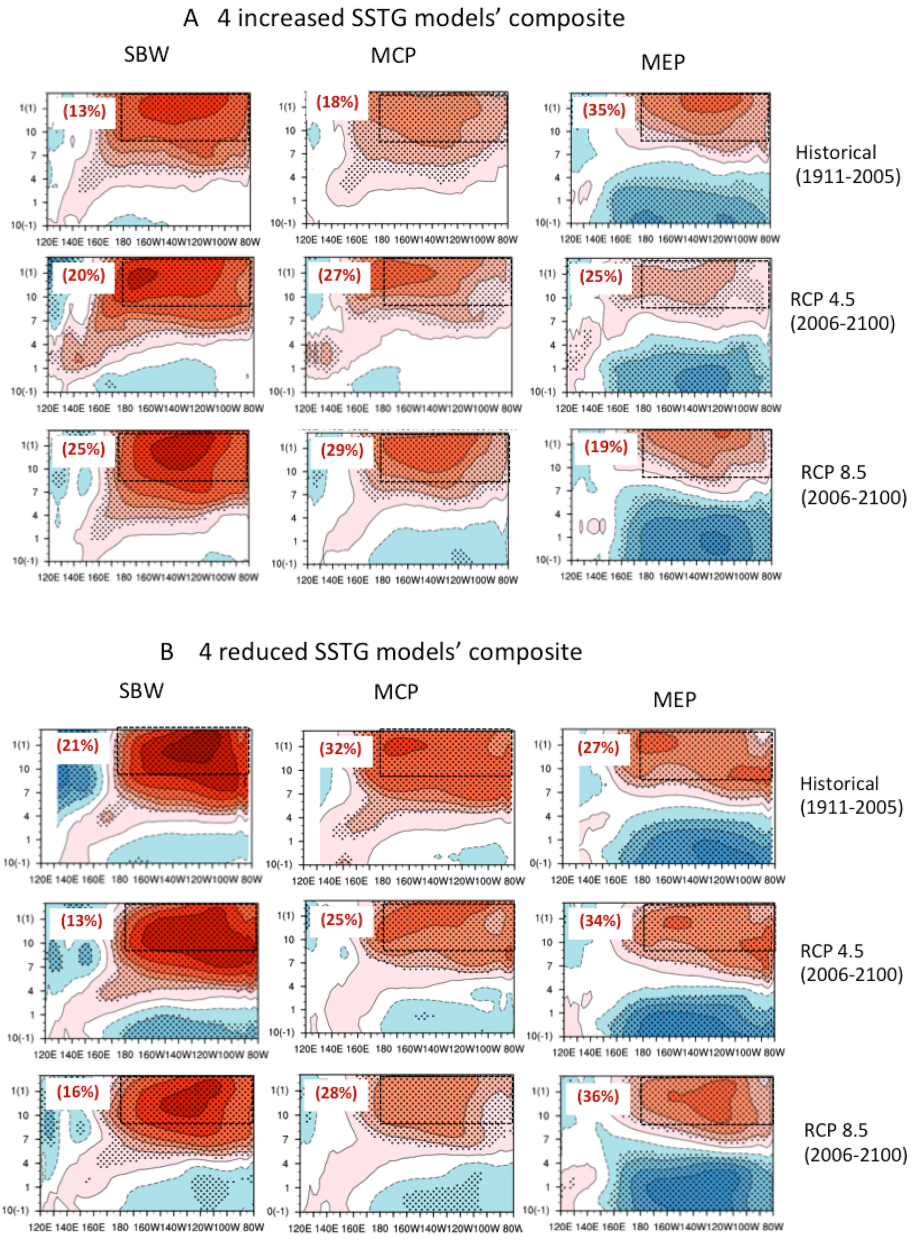

Figure S6 Composite evolutions of the equatorial Pacific SST anomalies in three types of El Niño onset, similar to those in Fig. 1 except derived from CMIP coupled models. (A) composite evolution from 4 CMIP5 models that project an increased zonal mean SST gradient (SSTG) in the RCP 4.5 and RCP 8.5 scenario. (B) composite evolution from 4 models that project a decreased zonal mean SST gradient in the RCP 4.5 and RCP 8.5 scenario. In (A) and (B), the upper, middle and lower panels compare the results from the corresponding historical (HIST), RCP 4.5 and RCP 8.5 runs, respectively. The left, central and right panels are for the SBW, MCP and MEP events, respectively. The percentages in the brackets denote the frequency of occurrence of each types of El Niño. The El Nino intensity is defined by the SST anomalies averaged in the dashed boxes (the equatorial region from the dateline to 80°W, and the time period from Sept (0) to March (1). To facilitate comparison, the period of integration is 95 yrs for both the historical run (1911-2005) and the RCP4.5 and RCP 8.5 run (2006-2100).

References

1. Rayner NA, *et al.* (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres* 108(D14):4407.

2. Huang B, *et al.* (2017) Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate* 30(20):8179-8205.

3. Carton JA & Giese BS (2008) A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA). *Monthly Weather Review* 136(8):2999-3017.

4. Giese BS & Ray S (2011) El Niño variability in simple ocean data assimilation (SODA), 1871–2008. *Journal of Geophysical Research: Oceans* 116(C2).

5. Saha S, *et al.* (2006) The NCEP climate forecast system. *Journal of Climate* 19(15):3483-3517.

6. Compo GP, *et al.* (2011) The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society* 137(654):1-28.

7. Kanamitsu M, *et al.* (2002) NCEP–DOE AMIP-II Reanalysis (R-2). *Bulletin of the American Meteorological Society* 83(11):1631-1644.

8. Poli P, *et al.* (2016) ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate* 29(11):4083-4097.

9. Uppala SM, *et al.* (2005) The ERA-40 re-analysis. *Quarterly Journal of the royal meteorological society* 131(612):2961-3012.

10. Dee DP, *et al.* (2011) The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society* 137(656):553-597.

11. Taylor KE, Stouffer RJ, & Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society* 93(4):485-498.

12. Schneider U, *et al.* (2014) GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology* 115(1):15-40.

13. Harris I, Jones P, Osborn T, & Lister D (2014) Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International Journal of Climatology* 34(3):623-642.

14. Wilks DS (2011) *Statistical methods in the atmospheric sciences* (Academic press).

15. Kaufman L & Rousseeuw PJ (2009) *Finding groups in data: an introduction to cluster analysis* (John Wiley & Sons).

16. An S-I & Jin F-F (2004) Nonlinearity and asymmetry of ENSO. *Journal of Climate* 17(12):2399-2412.

17. Lorbacher K, Dommenget D, Niiler P, & Köhl A (2006) Ocean mixed layer depth: A subsurface proxy of ocean‐atmosphere variability. *Journal of Geophysical Research: Oceans* 111(C7).