Supplementary Information

Supplementary Information 1

Descriptions of the CNN model, employed computer, and parameter settings

A CNN consists of multiple layers: an input layer, multiple hidden layers, and a single output layer. Information (data map) flows forward from the input layer, through the hidden layers and toward the output layer, which shows classification of the information. The building block of CNN is the artificial neuron. Each neuron receives one or more input signals, and calculates a weighted sum of input signals, and the resultant signal is transmitted to neurons in the next layer and so forth. Hidden layers of the CNN contain convolution layer(s), which is a learnable filter that activates when it detects some specific type of feature at some spatial position in the data map. Another important component of a CNN is the pooling layer(s), which reduces the data map through down-sampling.

We employed the LeNet CNN for this study. This CNN is composed of eight layers, one input, two sets of convolution and pooling layers, two fully connected layers, and one output layer. Intuitively, the convolution and pooling layers function as a feature extraction block, while fully connected layers function as a classification block. LeNet was originally developed for classifying handwritten digits (10 categories). Recent CNNs such as AlexNet and GoogLeNet contain a larger number of hidden layers and were designed for classifying data maps into tens of thousands of categories. Here, as we categorize VCEs into only 15 categories, LeNet provides be a reasonable option for balancing capability and computational load.

Besides the number of hidden layers, LeNet has some differences compared to other modern CNNs. In fully connected CNN layers, information is converted by an activating function before transmitting to the next layer. LeNet uses Sigmoid as an activation function, while other CNNs typically use the rectified linear unit (ReLU), which only transforms negative input values into zero. For reducing the size of data map in the pooling layers, LeNet uses the subsampling approach, while other modern CNNs use the max pooling approach, which selects the maximum values in the assigned local data-map. Therefore, other CNNs should be explored for increasing accuracy for training, although this will increase the computational load.

The computer employed to execute the learning had Ubuntu 16.04 LTS installed as the operating system and was equipped with an Intel Core i7-8700 CPU, 16 GB of RAM, and an NVIDIA GeForce GTX1080Ti graphics card, which accelerates the learning procedure. On the computer, the NVIDIA DIGITS 6.0.0 software (Caffe version: 0.15.14) served as the basis for CNN execution, and LeNet was employed to train the CNN via the TensorFlow library. Parameter settings remained set to the default values, except for the "% for validation" parameter, which determines the fraction of training data to be allocated for validation of the trained model. Since different climatic datasets are used for validation purposes as described in the main text, the default value of 25% for the parameter was changed to 0%, thereby allocating the entirety of the data to training.

30   Unchanged parameters include the training epochs, which determine how many passes through the training data are performed, and the base learning rate, which determines how quickly the CNN learns. In the training of a CNN, neuron weights and biases were updated and adjusted repeatedly by using the supervised learning back-propagation algorithm along with the Stochastic-Gradient-Descent solver, which is the optimization function for adjusting weights according to the error they caused. The amount that the weights are updated during training is referred to as the learning rate. Learning rates that

35   are too small result in lower training efficiency and may get stuck on a local-optimal solution. By contrast, learning rates that are too large result in a less optimal solution, may cause oscillation of model performance over training epochs, and may result in a numerical overflow in the worst case. Here, we employed the step down procedure for the learning rate: it starts from 0.01, and when 33% of training epochs pass, the learning rate is multiplied by 0.1 (in the DIGITs implementations, policy is "Step Down", base learning rate is 0.01, step size is 33%, and gamma is 0.1). As we conducted 30 training epochs,

40   the learning rate was 0.01 during the 1st to 10th epochs, 0.001 during 11th to 20th epochs, and 0.0001 for 21th to 30th epochs. For all training in this study, we visually confirmed that the model performance gradually increases through training epochs, and no oscillation pattern was observed.

Table S1

45 Global climate datasets used in this study.

Data ID	Description	Citation	URL
CRU TS v. 4.00	Observation-based climate dataset. Monthly, 0.5° grid. Period available: 1901–2015.	Harris et al. (2020)	http://www.cru.uea.ac.uk/cru/data/hrg/
NCEP/NCAR reanalysis	Climate dataset incorporating observations and numerical weather prediction model output. Daily, T62 (latitude 94 × longitude 192). Period available: 1948–present.	Kalnay et al. (1996)	https://rda.ucar.edu/datasets/ds090.0/
HadGEM2-ES	Output of the Hadley Centre Global Environment Model version 2. Monthly, latitude 1.25 degree × longitude 1.875 degree. Period available: 1861–2005 (historical) and 2006–2100 (future scenarios).	Collins et al. (2011)	https://esgf-node.llnl.gov/projects/cmip5/
MIROC-ESM	Output of the MIROC Earth System Model version 2010. Monthly, T42 (latitude 64 × longitude 128). Period available: 1861–2005 (historical) and 2006–2100 (future scenarios).	Watanabe et al. (2011)	https://esgf-node.llnl.gov/projects/cmip5/

Reference

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire,
50 A., and Woodward, S.: Development and evaluation of an Earth-System model-HadGEM2, Geoscientific Model Development, 4, 1051-1075, 10.5194/gmd-4-1051-2011, 2011.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Sci Data, 7, 109, 10.1038/s41597-020-0453-3, 2020.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G.,
55 Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437-471, 1996.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., and Kawamiya, M.: MIROC-ESM
60 2010: model description and basic results of CMIP5-20c3m experiments, Geoscientific Model Development, 4, 845-872, 10.5194/gmd-4-845-2011, 2011.

Table S2

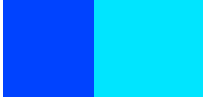
Influence of VCEs on training and test accuracies for simulation of the geographical distribution of terrestrial biomes using a selection of different climate datasets. The values in brackets indicate the computational time for training. For each climate dataset, the annual mean bio-temperature and the annual precipitation from 1971 to 1980 were standardized and log transformed, then employed for drawing VCEs.

Examples of VCE, representing mean annual bio- temperature and precipitation.	Training accuracy	Test accuracy for each climatic data set			
		NCEP/ NCAR	HadGEM2 -ES	MIROC -ESM	Average
A RGB colour tile. Red and green ranges were used to represent temperature and precipitation, respectively [15 minutes]	58.3%	45.6%	48.6%	41.3%	48.5%
Two tiles using the "Topo.colour" colour scheme of R [34 minutes]	59.1%	45.2%	48.5%	41.6%	48.6%
Two tiles using the "Gray.scale" colour scheme of R [23 minutes]	58.5%	45.4%	48.6%	41.8%	48.6%
Two pie charts [22 minutes]	59.7%	45.2%	47.9%	41. 4%	50.9%

Table S3

70 Design of combination experiment of annual means of climatic variables and its resulting accuracies. "Air Temp" is the annual average of bio-temperature. "Prec." is the annual precipitation. "Rad Short" is the annual average of downward shortwave radiation (whole day mean). "Humid" is the annual average of the volumetric humidity of air. For models with one climate variable, the selected climatic variable was allocated to the R channel. For models with two climate variables, the R and G channels were used. For models with three climate variables, all RGB channels were used.

Model no.	Combinations of climate variables (annual means)				Training accuracy	Test accuracy for each climatic data set			
	Air Temp	Prec	Rad Short	Humid		NCEP/ NCAR	HadGEM2 -ES	MIROC -ESM	Average
1	○				38.1%	34.3%	34.7%	31.5%	33.5%
2		○			35.2%	29.5%	31.8%	28.9%	30.1%
3			○		35.8%	30.9%	23.5%	21.0%	25.1%
4				○	36.3%	29,4%	33.2%	25.7%	29.5%
5	○	○			57.3%	45.6%	48.8%	41.4%	45.3%
6	○		○		48.7%	39.9%	33.3%	30.8%	34.7%
7	○			○	54.3%	41.4%	45.6%	40.7%	42.6%
8		○	○		52.1%	39.8%	40.5%	36.5%	38.9%
9		○		○	52.3%	42.2%	46.5%	35.4%	41.4%
10			○	○	50.9%	38.4%	31.4%	27.6%	32.5%
11	○	○	○		63.0%	46.4%	48.2%	44.6%	46.4%
12	○	○		○	62.5%	45.1%	49.6%	42.2%	45.6%
13	○		○	○	58.5%	41.5%	38.9%	35.3%	38.6%
14		○	○	○	61.3%	42.0%	41.9%	36.6%	40.2%

75

Table S 4

Design of combination experiment of monthly means of climatic variables and its resulting accuracies. "Air Temp" is the

Model no.	Combinations of climatic variables (monthly means)				Training accuracy	Test accuracy for each climate data set			
	Air Temp	Prec	Rad Short	Humid		NCEP/ NCAR	HadGEM2 -ES	MIROC -ESM	Average
1	○				61.8%	51.3%	51.3%	39.2%	47.3%
2		○			79.4%	41.7%	36.8%	33.9%	37.5%
3			○		61.2%	39.5%	29.7%	31.3%	33.5%
4				○	68.3%	39.7%	40.7%	34.8%	38.4%
5	○	○			83.2%	52.7%	51.5%	43.5%	49.2%
6	○		○		72.2%	51.8%	39.8%	36.9%	42.8%
7	○			○	73.7%	52.4%	49.4%	41.8%	47.9%
8		○	○		80.5%	48.3%	43.4%	36.3%	42.7%
9		○		○	81.5%	47.4%	48.7%	40.1%	45.4%
10			○	○	75.0%	48.4%	43.0%	38.6%	43.3%
11	○	○	○		82.4%	52.1%	51.5%	42.6%	48.7%
12	○	○		○	81.8%	51.8%	51.3%	41.2%	48.1%
13	○		○	○	77.9%	54.9%	44.4%	42.9%	47.4%
14		○	○	○	82.6%	49.7%	51.0%	41.8%	47.5%

monthly average of the air temperature. Other climatic items are as described in the caption in table S3.

80

Table S5

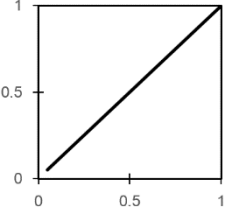
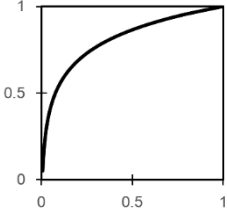
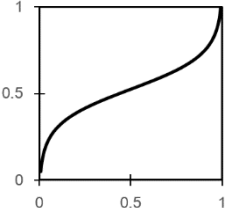
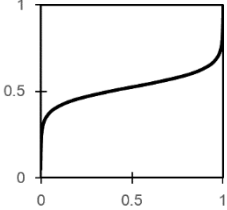
Influence of transformation of climatic variables for drawing VCE on training and test accuracies of the biome mapping
simulations. Before drawing the VCE, each climatic item was standardized to between 0.0 and 1.0. These standardized
climatic variables were then transformed by one of the four transformation schemes for each experiment. Graphs on the
leftmost column show how the original variables (horizontal axis) were transformed. This experiment employs VCE with a
RGB colour tile, in which the red and green ranges represent the annual mean bio-temperature and annual precipitation,
respectively.

Scheme to transform climatic variables for drawing VCE	Training accuracy	Test accuracy for each climatic data set			
		NCEP/ NCAR	HadGEM2 -ES	MIROC -ESM	Average
Linear	57.8%	45.5%	41.3%	48.6%	45.1%
Log	56.9%	45.9%	41.6%	48.2%	45.2%
Sigmoid (gain=5)	56.9%	45.3%	40.9%	48.5%	44.9%
Sigmoid (gain=10)	56.2%	44.5%	41.0%	48.3%	44.6%

Table S6

Design of how the monthly mean air temperature and monthly precipitation were superimposed to colour channels for drawing VCE, and its resulting accuracies. "R", "G", and "B", respectively, represent the red, green, and blue channels, of the RGB colour representation. Although there are no major differences in the models of either the training or test accuracies, we selected model no. 4, which has the highest average test accuracy.

Model no.	Colour channel		Training accuracy	Test accuracy for each data set			
	Air temp	Prec		NCEP/ NCAR	HadGEM2 -ES	MIROC -ESM	Average
1	R	G	81.2%	53.2%	52.2%	42.7%	49.4%
2	R	B	81.8%	52.4%	52.3%	42.8%	49.2%
3	G	R	81.3%	53.2%	50.5%	43.2%	49.0%
4	B	R	81.4%	52.9%	52.4%	43.6%	49.6%
5	G	B	82.7%	53.6%	52.3%	42.3%	49.4%
6	B	G	81.4%	52.7%	51.0%	44.0%	49.3%

Table S7

Confusion matrix for biome classification for training accuracy with annual mean bio-temperature and annual precipitation. Columns represent the true classes, while rows represent the CNN model prediction. The shaded cells along the upper-left to lower-right diagonal represent correct classifications. For each cell, the upper line indicates the number of simulation grids, while the lower line indicates its percentage within the column, which in turn indicates the fraction of correct classifications of the corresponding biome class.

	Predicted class														
Actual class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Polar Desert/Rock/Ice	324	0	463	0	0	0	0	88	1	29	0	0	0	0	0
	35.8%	0.0%	51.2%	0.0%	0.0%	0.0%	0.0%	9.7%	0.1%	3.2%	0.0%	0.0%	0.0%	0.0%	0.0%
2. Desert	1	4689	0	650	23	145	7	20	4	12	4	0	0	0	0
	0.0%	84.4%	0.0%	11.7%	0.4%	2.6%	0.1%	0.4%	0.1%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%
3. Tundra	48	5	3192	9	0	50	2	1386	8	157	8	8	0	0	0
	1.0%	0.1%	65.5%	0.2%	0.0%	1.0%	0.0%	28.4%	0.2%	3.2%	0.2%	0.2%	0.0%	0.0%	0.0%
4. Open Shrubland	0	637	0	2383	119	901	391	69	3	28	27	13	0	0	4
	0.0%	13.9%	0.0%	52.1%	2.6%	19.7%	8.5%	1.5%	0.1%	0.6%	0.6%	0.3%	0.0%	0.0%	0.1%
5. Dense Shrubland	0	46	0	640	325	205	719	27	0	1	123	4	0	2	58
	0.0%	2.1%	0.0%	29.8%	15.1%	9.5%	33.4%	1.3%	0.0%	0.0%	5.7%	0.2%	0.0%	0.1%	2.7%
6. Grassland/Steppe	0	60	8	518	213	3190	561	520	99	140	446	128	0	12	127
	0.0%	1.0%	0.1%	8.6%	3.5%	53.0%	9.3%	8.6%	1.6%	2.3%	7.4%	2.1%	0.0%	0.2%	2.1%
7. Savanna	0	8	8	380	145	487	3680	426	18	59	542	162	0	47	932
	0.0%	0.1%	0.1%	5.5%	2.1%	7.1%	53.4%	6.2%	0.3%	0.9%	7.9%	2.3%	0.0%	0.7%	13.5%
8. Eg/Dc Mixed Forest	2	1	935	16	34	127	124	5750	77	940	724	317	0	11	116
	0.0%	0.0%	10.2%	0.2%	0.4%	1.4%	1.4%	62.7%	0.8%	10.2%	7.9%	3.5%	0.0%	0.1%	1.3%
9. Bor Dc Forest	0	0	5	0	0	0	0	803	334	219	2	2	0	0	0
	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.0%	58.8%	24.5%	16.0%	0.1%	0.1%	0.0%	0.0%	0.0%
10. Bor Eg Forest	2	0	28	0	0	18	0	1406	163	2160	5	23	0	0	0
	0.1%	0.0%	0.7%	0.0%	0.0%	0.5%	0.0%	37.0%	4.3%	56.8%	0.1%	0.6%	0.0%	0.0%	0.0%
11. Tmp Dc Forest	0	0	0	7	6	182	17	203	0	5	1698	132	0	0	2
	0.0%	0.0%	0.0%	0.3%	0.3%	8.1%	0.8%	9.0%	0.0%	0.2%	75.4%	5.9%	0.0%	0.0%	0.1%
12. Tmp Nl Eg-Forest	0	0	0	0	8	90	61	335	0	29	372	652	1	12	111
	0.0%	0.0%	0.0%	0.0%	0.5%	5.4%	3.7%	20.0%	0.0%	1.7%	22.3%	39.0%	0.1%	0.7%	6.6%
13. Tmp Bl-Eg Forest	0	0	0	1	20	22	53	38	0	1	208	67	2	2	35
	0.0%	0.0%	0.0%	0.2%	4.5%	4.9%	11.8%	8.5%	0.0%	0.2%	46.3%	14.9%	0.4%	0.4%	7.8%
14. Tro Dc Forest	0	0	0	25	2	8	1397	17	1	3	41	16	0	71	424
	0.0%	0.0%	0.0%	1.2%	0.1%	0.4%	69.7%	0.8%	0.0%	0.1%	2.0%	0.8%	0.0%	3.5%	21.1%
15. Tro Eg Forest	0	0	0	2	5	4	800	17	0	0	60	61	0	32	4588
	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	14.4%	0.3%	0.0%	0.0%	1.1%	1.1%	0.0%	0.6%	82.4%

Percentage of correct classification for all grids: 57.7%.

Abbreviations: Bor, boreal; Tmp, temperate; Tro, tropical; Eg, evergreen; Dc, deciduous; Nl, needleleaf; Bl, broadleaf.

Table S8

Same as table S7, but confusion matrix for biome classification for training accuracy with monthly mean air temperature and monthly precipitation.

Predicted class

Actual class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Polar Desert/Rock/Ice	574	7	298	3	0	1	0	19	0	3	0	0	0	0	0
	63.4%	0.8%	32.9%	0.3%	0.0%	0.1%	0.0%	2.1%	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%
2. Desert	5	5048	10	367	46	65	8	2	0	0	0	0	1	3	0
	0.1%	90.9%	0.2%	6.6%	0.8%	1.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%
3. Tundra	104	2	4059	1	0	77	10	533	0	71	8	2	0	4	2
	2.1%	0.0%	83.3%	0.0%	0.0%	1.6%	0.2%	10.9%	0.0%	1.5%	0.2%	0.0%	0.0%	0.1%	0.0%
4. Open Shrubland	3	626	11	2778	226	486	333	1	0	13	10	31	0	36	21
	0.1%	13.7%	0.2%	60.7%	4.9%	10.6%	7.3%	0.0%	0.0%	0.3%	0.2%	0.7%	0.0%	0.8%	0.5%
5. Dense Shrubland	0	24	1	415	955	201	296	34	0	1	60	7	8	77	71
	0.0%	1.1%	0.0%	19.3%	44.4%	9.3%	13.8%	1.6%	0.0%	0.0%	2.8%	0.3%	0.4%	3.6%	3.3%
6. Grassland/Steppe	1	22	45	315	99	4426	521	211	11	33	96	69	16	35	122
	0.0%	0.4%	0.7%	5.2%	1.6%	73.5%	8.7%	3.5%	0.2%	0.5%	1.6%	1.1%	0.3%	0.6%	2.0%
7. Savanna	0	11	4	241	174	485	4314	305	62	13	184	63	61	323	654
	0.0%	0.2%	0.1%	3.5%	2.5%	7.0%	62.6%	4.4%	0.9%	0.2%	2.7%	0.9%	0.9%	4.7%	9.5%
8. Eg/Dc Mixed Forest	0	1	660	10	29	76	175	6700	255	611	428	96	20	17	96
	0.0%	0.0%	7.2%	0.1%	0.3%	0.8%	1.9%	73.0%	2.8%	6.7%	4.7%	1.0%	0.2%	0.2%	1.0%
9. Bor Dc Forest	0	0	2	0	0	9	25	208	1079	36	6	0	0	0	0
	0.0%	0.0%	0.1%	0.0%	0.0%	0.7%	1.8%	15.2%	79.0%	2.6%	0.4%	0.0%	0.0%	0.0%	0.0%
10. Bor Eg Forest	0	0	35	0	0	8	18	758	18	2935	2	31	0	0	0
	0.0%	0.0%	0.9%	0.0%	0.0%	0.2%	0.5%	19.9%	0.5%	77.1%	0.1%	0.8%	0.0%	0.0%	0.0%
11. Tmp Dc Forest	0	2	1	4	5	48	33	137	1	8	1828	127	58	0	0
	0.0%	0.1%	0.0%	0.2%	0.2%	2.1%	1.5%	6.1%	0.0%	0.4%	81.2%	5.6%	2.6%	0.0%	0.0%
12. Tmp Nl Eg-Forest	0	0	0	21	21	26	49	179	0	165	229	909	53	3	16
	0.0%	0.0%	0.0%	1.3%	1.3%	1.6%	2.9%	10.7%	0.0%	9.9%	13.7%	54.4%	3.2%	0.2%	1.0%
13. Tmp Bl-Eg Forest	0	0	0	4	28	54	31	19	0	0	44	1	246	3	19
	0.0%	0.0%	0.0%	0.9%	6.2%	12.0%	6.9%	4.2%	0.0%	0.0%	9.8%	0.2%	54.8%	0.7%	4.2%
14. Tro Dc Forest	0	0	1	5	20	18	344	18	0	0	2	0	1	1241	355
	0.0%	0.0%	0.0%	0.2%	1.0%	0.9%	17.2%	0.9%	0.0%	0.0%	0.1%	0.0%	0.0%	61.9%	17.7%
15. Tro Eg Forest	0	0	0	0	10	47	262	32	0	0	0	1	2	151	5064
	0.0%	0.0%	0.0%	0.0%	0.2%	0.8%	4.7%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	2.7%	90.9%

Percentage of correct classification for all grids: 73.6%.
Abbreviations: Bor, boreal; Tmp, temperate; Tro, tropical; Eg, evergreen; Dc, deciduous; Nl, needleleaf; Bl, broadleaf.

Table S9

CNN model accuracies for biome distribution simulations. These accuracies were obtained using the model trained by the climatic dataset on the row, with the climate dataset on the column as an input reconstruction. Therefore, the shaded cells show the accuracy when the climate datasets for training and reconstruction were identical. For each climate dataset, the monthly mean-temperature and monthly precipitation during 1971 to 1980 were standardized and log transformed, then used for drawing the RGB-colour tile VCEs.

	CRU	NCEP/NCAR	Miroc-ESM	HadGEM2-ES
CRU	0.736	0.559	0.478	0.512
NCEP/NCAR	0.553	0.704	0.431	0.485
Miroc-ESM	0.540	0.394	0.701	0.417
HadGEM2-ES	0.430	0.505	0.450	0.712

Table S10

Results of the sensitivity test in which training and test accuracy were compared among models trained and validated using monthly climate averaged over one of three time periods.

Time period	Training accuracy	Test accuracy			
		NCEP/ NCAR	HadGEM2 -ES	MIROC -ESM	Average
10 years (1971-1980; Control)	81.4%	52.9%	52.4%	43.6%	49.6%
20 years (1961-1980)	77.5%	52.2%	52.4%	44.8%	49.8%
30 years (1951-1980)	76.0%	52.4%	52.7%	43.7%	49.6%

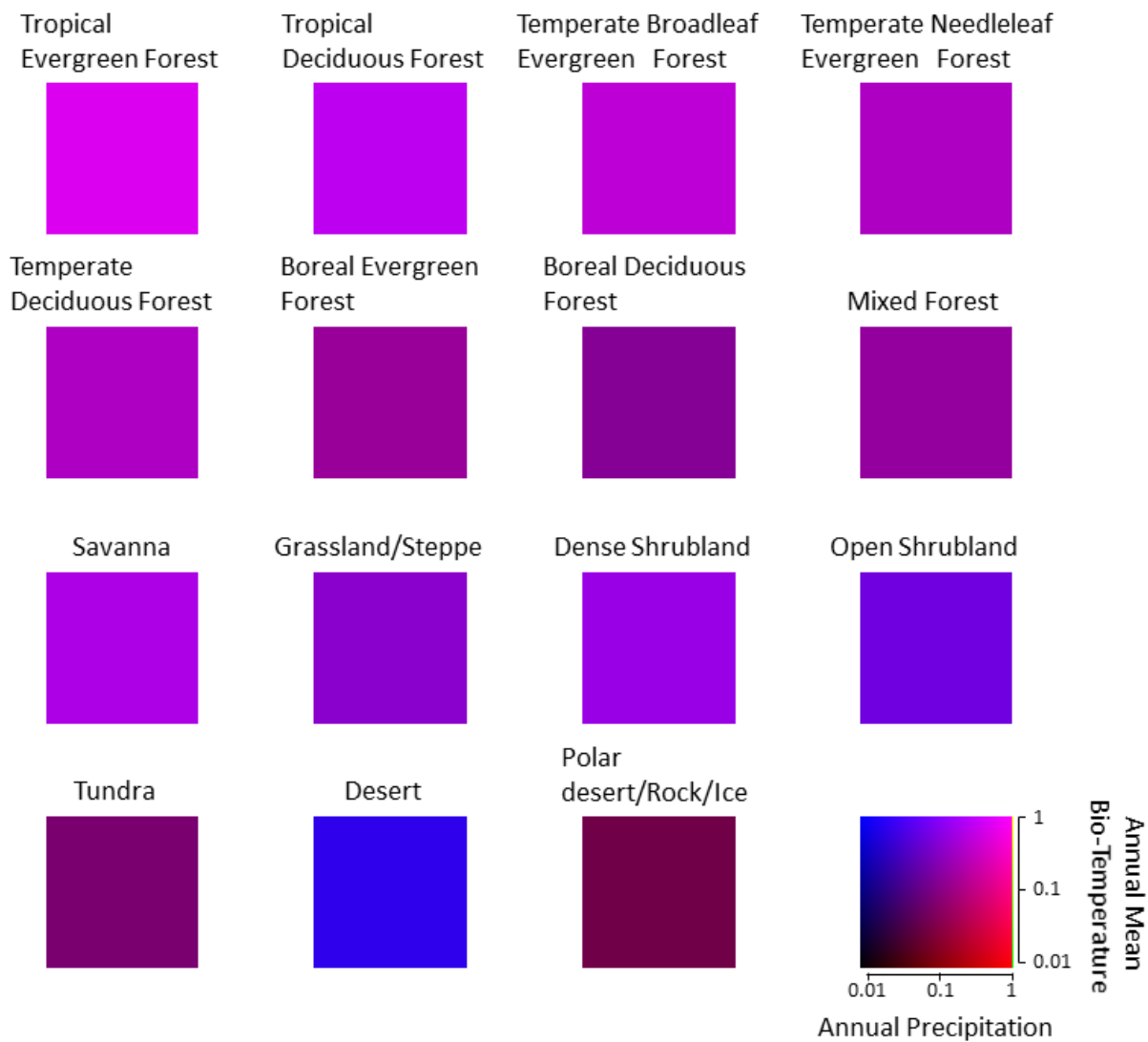
Table S11

Results of the sensitivity test in which training and test accuracy were compared among models trained using one of four aggregation grains of training data grids. Test accuracy for the CRU dataset in this analysis was obtained by model evaluation for grains not used in training.

Aggregation grain of CRU data grids	Training accuracy	Test accuracy			
		CRU	NCEP/ NCAR	HadGEM2 -ES	MIROC -ESM
1×1 (0.5°)	80.2%	80.4%	53.1%	50.5%	43.5%
2×2 (1.0°)	80.6%	78.2%	53.3%	51.4%	42.9%
4×4 (2.0°)	81.2%	76.1%	52.8%	51.7%	43.9%
8×8 (4.0°)	81.0%	72.2%	52.7%	51.2%	42.6%

Figure S1

130   Examples of training images representing annual mean bio-temperature and annual precipitation of the CRU dataset. Climatic conditions averaged over 1971 to 1980 for each biome type were employed. The bottom-right image shows how the panel colour was determined by standardized climatic variables.

Figure S2

Examples of training images representing monthly mean air temperature and monthly precipitation from the CRU dataset. Climatic conditions, averaged from 1971 to 1980, for each biome type were employed. For the Northern hemisphere, the leftmost pillar represents the January climate, while rightmost pillar represents the December climate for each image. For the Southern hemisphere, the leftmost pillar represents the July climate, while the rightmost pillar represents the June climate for each image. The image at the bottom-right corner shows how the pillar colour was determined from the standardized climatic variables.

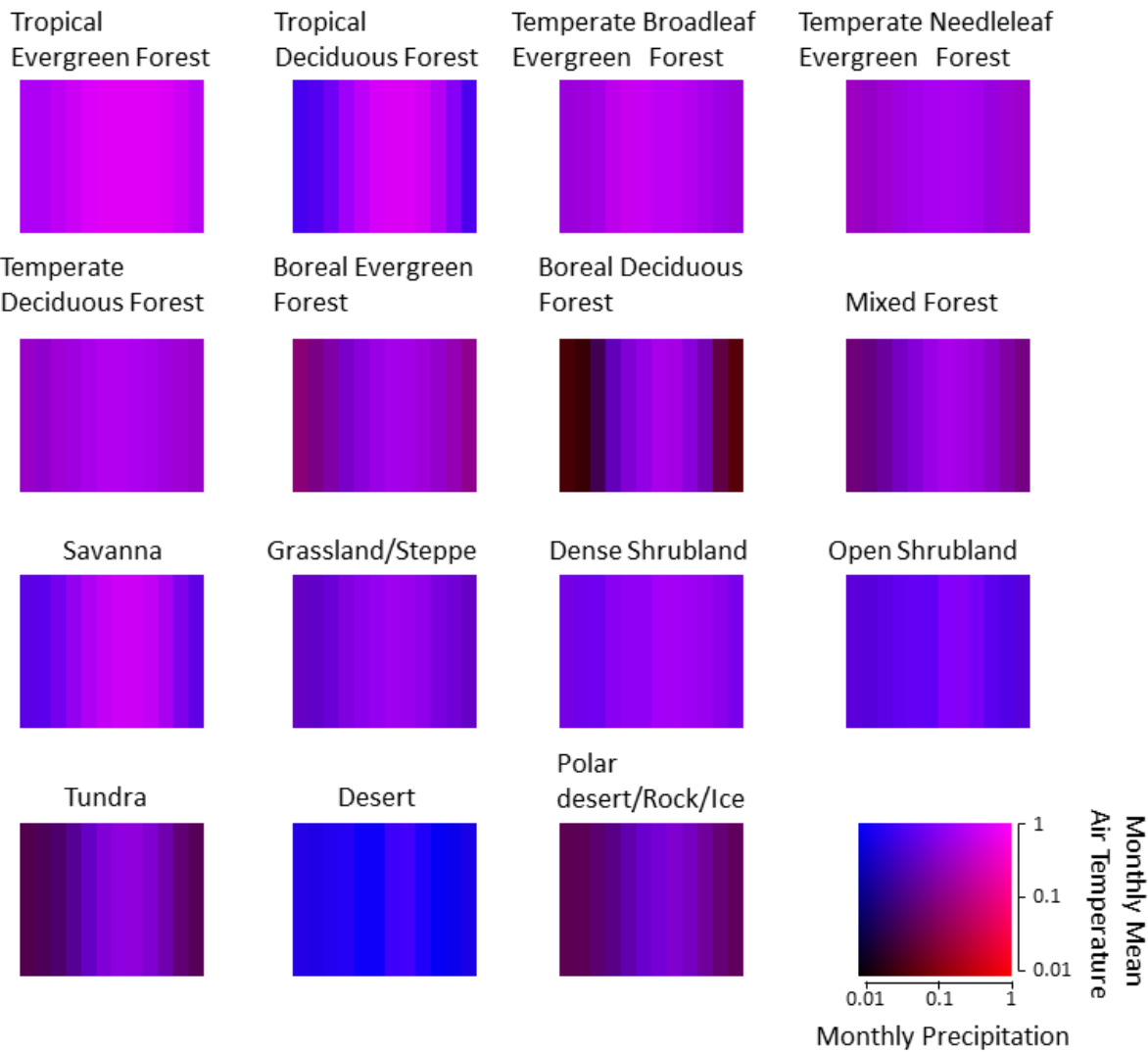



Figure S3

145 Geographical distribution of certainty of the most plausible biome from the CNN model that was trained with either annual (a, c, e, g) or monthly (b, d, f, h) mean climate images from the CRU dataset. Four climatic datasets were employed to generate these maps. (a, b) CRU dataset; (c, d) NCEP/NCAR dataset; (e, f) output of the Had2GEM-ES dataset; (g, h) output of the MIROC-ESM dataset.

150

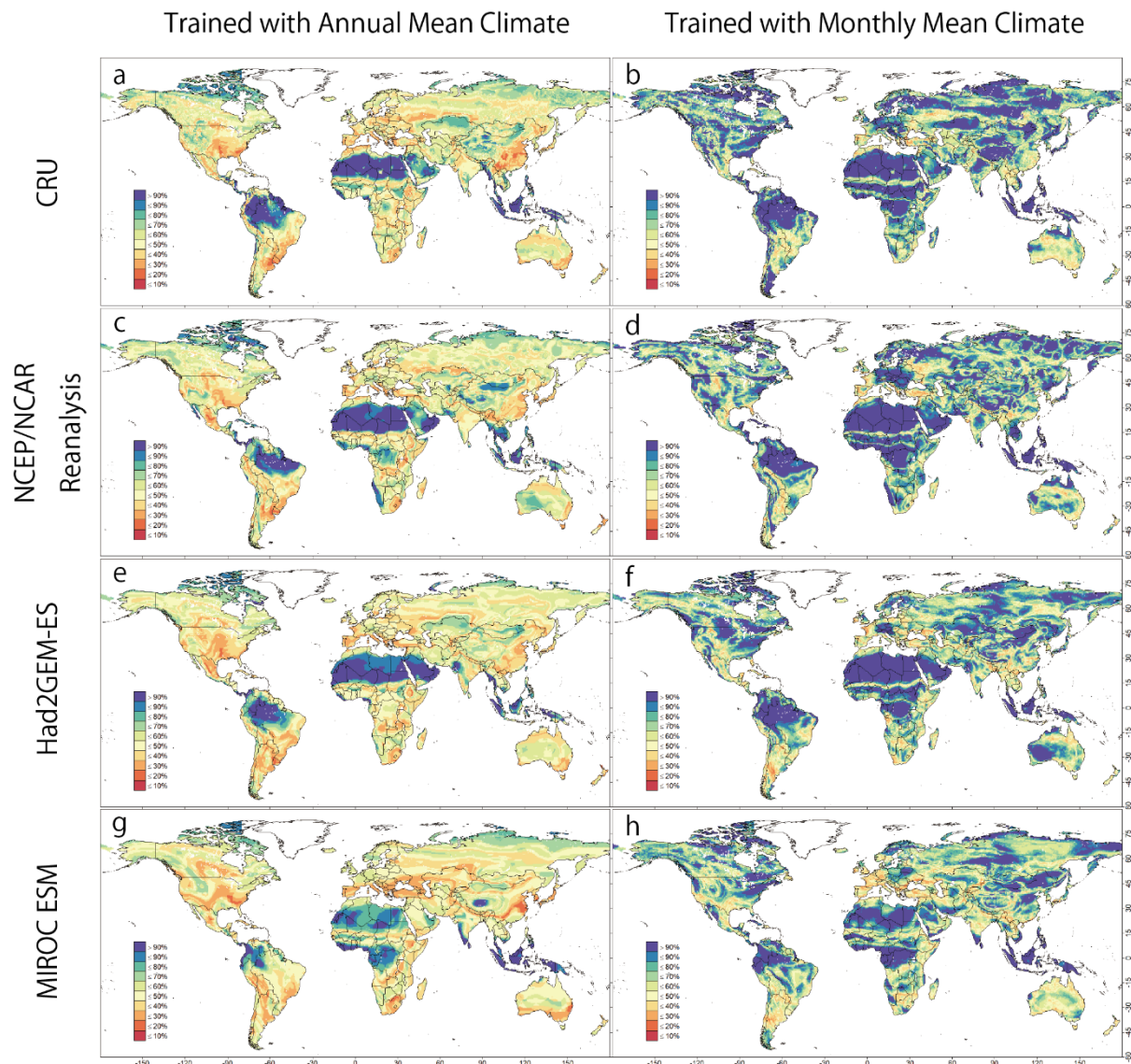



Figure S4

(a) Annual mean air temperature and (b) annual precipitation averaged from the 1971–1980 period of the CRU dataset. (c) Differences in annual mean air temperature (1971–1980 average) and (d) annual precipitation (1971–1980 average) between the CRU dataset and the NCEP/NCAR reanalysis data. The CRU and HadGEM2-ES dataset outputs for mean air temperature and annual precipitation are presented in (e) and (f), respectively. Similarly, the CRU and MIROC-ESM dataset outputs for mean air temperature and annual precipitation are presented in (g) and (h), respectively.

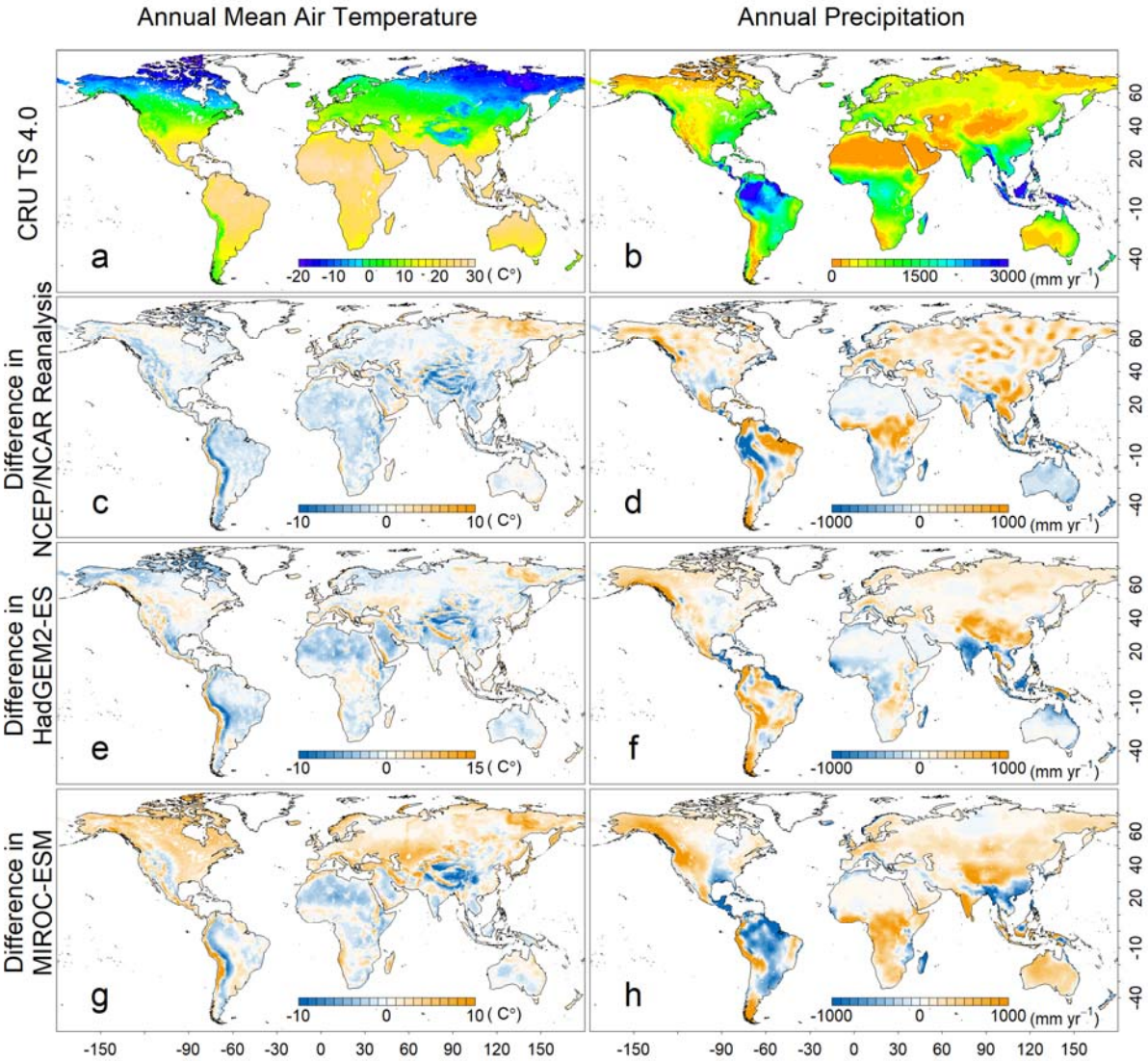



Figure S5

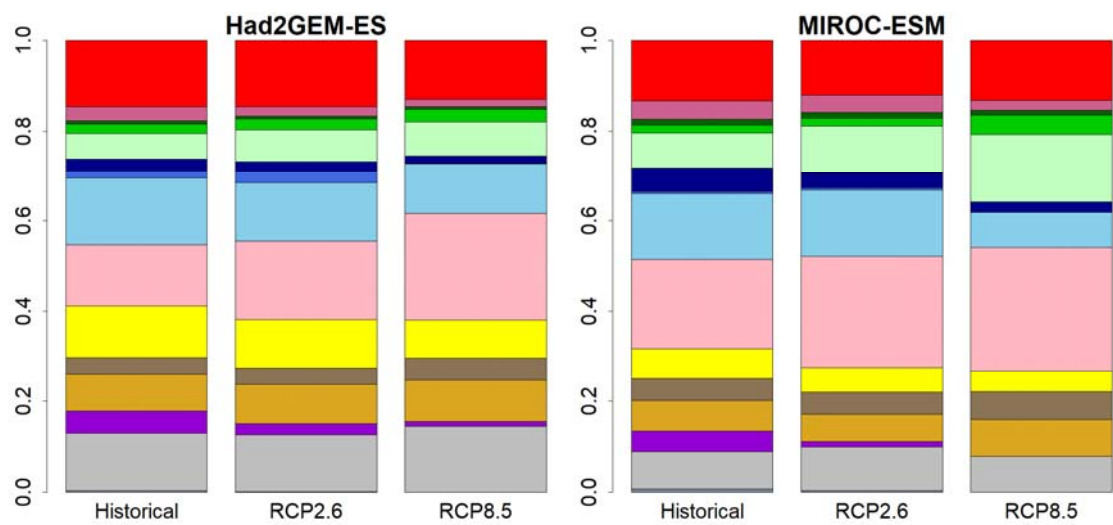
160   Global biome compositions computed by the CNN model using monthly means of two climate models (Had2GEM-ES and MIROC-ESM) under historical (1971–1980) and future (2091–2100) climatic conditions. Future climatic conditions were further divided into two scenarios (RCP2.6 and RCP8.5). Differences in grid areas along the latitude were taken into consideration. Colour definitions are available in figure 1.

165

Figure S6

(a) Mean annual air temperature and (b) annual precipitation averaged from the 1971–1980 period of the HadGEM2-EM dataset. Changes in (c) mean annual air temperature and (d) precipitation from 1971–1980 to 2091–2100 under the RCP2.6 scenario. Changes in (e) mean annual air temperature and (f) precipitation from 1971–1980 to 2091–2100 under the RCP8.5 scenario.

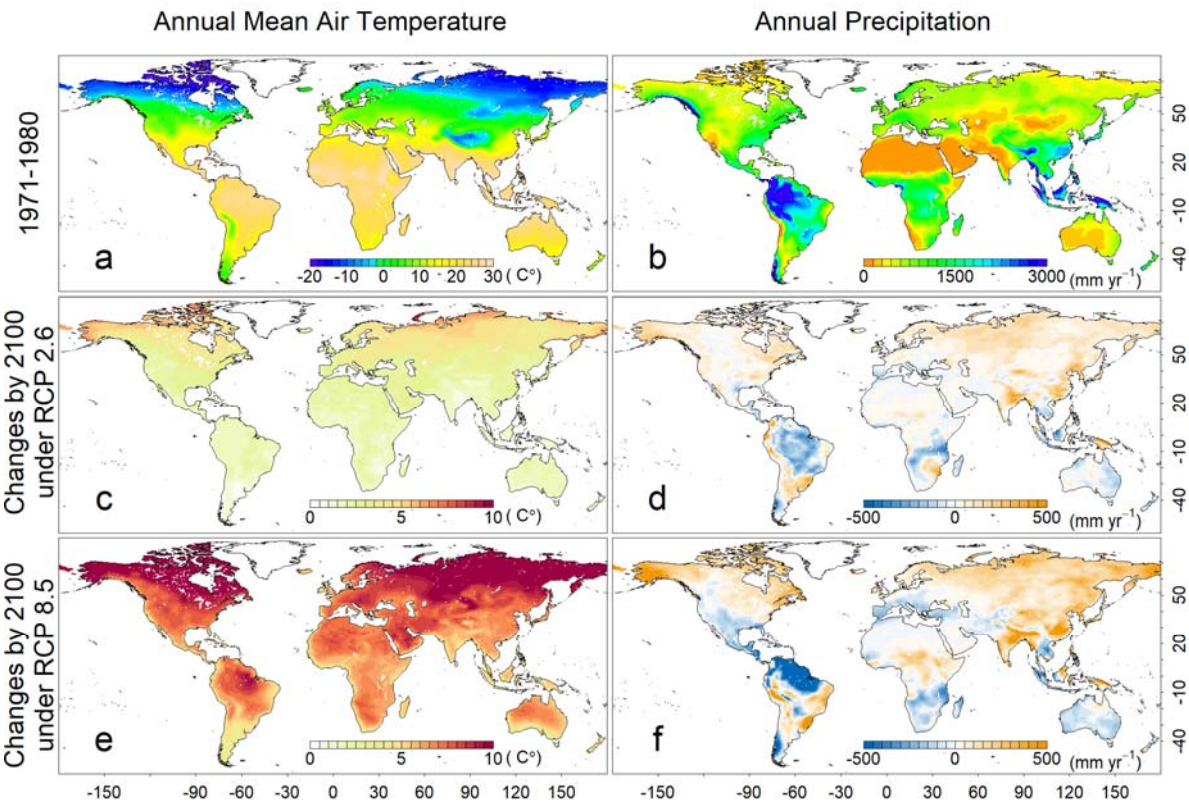



Figure S7

Same as figure S6, but with the MIROC-ESM dataset.

175

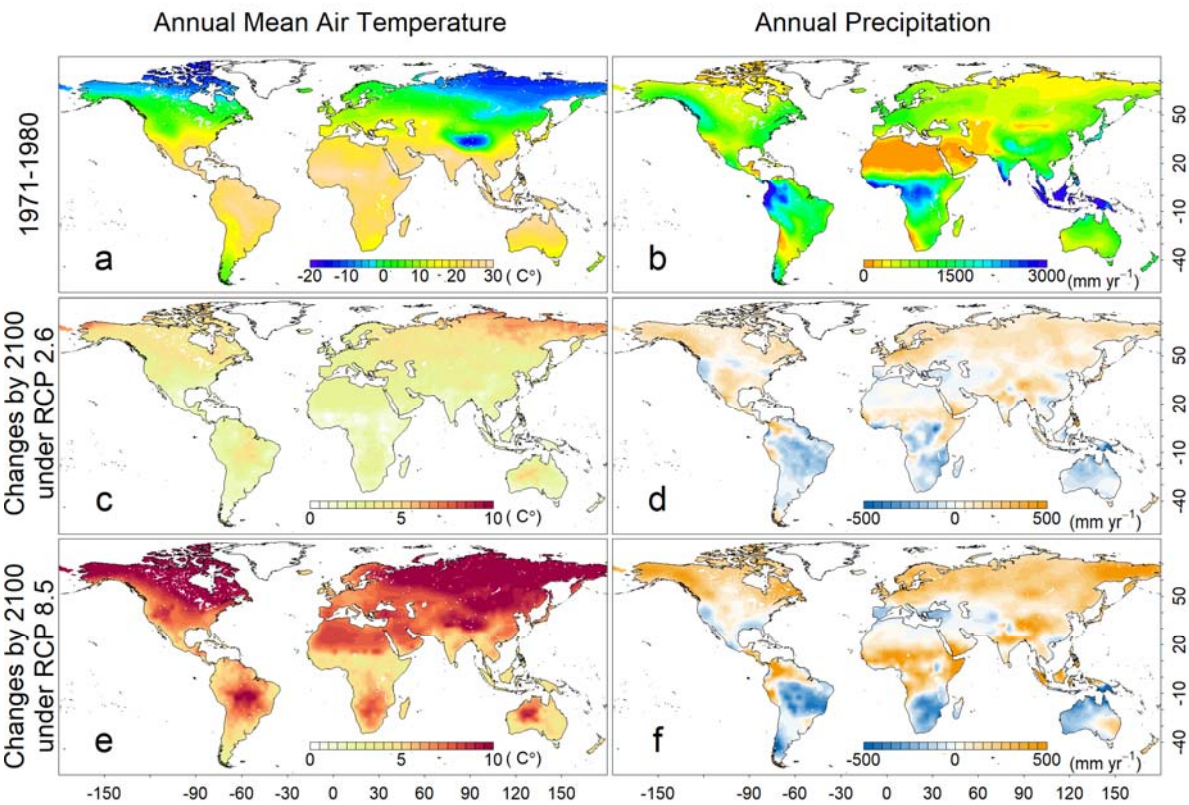



Figure S8

Same as figure 4, but certainty map of the predicted biome.