1 Supplements

This document supplements the paper “Temporal changes in size distributions of the Southern Ocean diatom *Fragilariopsis kerguelensis* through high-throughput microscopy of sediment trap samples” by Michael Kloster, Andrés S. Rigual-Hernández, Leanne K. Armand, Gerhard Kauer, Thomas W. Trull & Bánk Beszteri. Diatom Research, DOI: 10.1080/0269249X.2019.1626770.

1.1 Data

The raw data (morphometric measurements as well as the underlying light-microscopic images) are provided at PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.904325).

1.2 SHERPA settings

The configuration files are provided at PANGAEA. Settings for the low resolution scan can be found in “54S-800m.20x.dmini”, for the high resolution scan in “54S-800m.63x.dmini”

1.3 Modified SHERPA procedure

We adapted our established analysis workflow to compensate for the low imaging contrast of the mounted valves: Increased manual intervention was exerted for manual post-editing, as well as reviewing segmentation results by also addressing results of lower quality than described in (Kloster et al. 2017). “Quality” in this context relates to the question if a) an object is of interest, and b) the image processing and segmentation worked out well. SHERPA supports the user in tackling this problem by calculating the “ranking index” (RI), which increases with declining result quality.

Accordingly, for each slide, when selecting valves of interest in the low-resolution (20x magnification) scans, results from RIs of 0 to 6, respectively for sparse samples up to RI 10, were reviewed. If ca. 100 consecutive results of one particular RI did not represent suitable *F. kerguelensis* valves, the remaining results of this RI were skipped, because the probability that they contained further useful *F. kerguelensis* valve images was very low, and the procedure was resumed with results of the following RI. From the high-resolution scans (63x magnification), for each slide results were reviewed, reworked as needed, and *F. kerguelensis* valves selected for RIs 0 to at least 2. If the sample size was not sufficient (mostly below 300), results of higher RIs were added. The same applied if results of the highest RI changed the valve length distribution significantly, which was checked by a Kolmogorov-Smirnov test (see supplement “KS test results regarding RI”). In some cases, duplicate slides were scanned, to obtain higher sampling sizes.

Costae distance measurement is an experimental SHERPA feature and, until now, was tested on *Fragilariopsis* valves only (Beszteri et al. 2018). This frequency based method is mostly robust towards segmentation flaws, nevertheless results of costae distance ≥ 5 µm were assumed to be incorrect and thus rejected, which applied to ca. 0.5 % of all data. The valve image was segmented by Adaptive Thresholding (Bradski & Kaehler 2008), resulting in costae resp. valve border colored white and striae resp. image background colored black. The segmented image was sampled along the central 80 % of the valve’s major (apical) axis by the Bresenham algorithm (Bresenham 1965),. Along this sampled line, each white section was replaced by a 5 pixel wide binomial-smoothed peak, resulting in an artificial 2d-image of costae positions. This was transformed into the frequency domain by discrete Fourier transform (DFT). Average costae distance was calculated from the frequency of the maximum magnitude of the Fourier transformed.

1.4 Valves within the initial cell size range

Table S1. Abundance, percentage and relative frequency of valves within the initial cell size range

Cup	Sample size *1	n of valves within the initial cell size range (≥ 75.5 µm valve length) *1	% of valves within the initial cell size range (≥ 75.5 µm valve length) *1	Rel. frequency of valves within the initial cell size range (RFI) *2
1.01	300	2	0.67	0.06
1.02	569	1	0.18	0.03
1.03	525	1	0.19	0.04
1.04	178	0	0.00	0.00
1.05	455	0	0.00	0.07
1.06	1630	1	0.06	0.02
1.07	574	1	0.17	0.14
1.08	736	1	0.14	0.02
1.09	1010	3	0.30	0.04
1.10	907	1	0.11	0.03
1.11	338	0	0.00	0.00
1.11	338	0	0.00	0.00
1.12	409	0	0.00	0.00
1.13	478	2	0.42	0.02
1.13	478	2	0.42	0.00
1.14	420	0	0.00	0.03
1.15	379	2	0.53	0.05
1.16	145	0	0.00	0.00
1.17	793	5	0.63	0.12
1.18	592	3	0.51	0.18
1.19	955	4	0.42	0.15
1.20	811	4	0.49	0.28
2.01	160	2	1.25	0.14
2.02	327	5	1.53	0.18
2.03	430	5	1.16	0.04
2.04	600	0	0.00	0.01
2.05	467	4	0.86	0.09
2.06	718	0	0.00	0.03
2.07	1639	5	0.31	0.40

2.08	1360	9	0.66	0.19
2.09	1228	4	0.33	0.18
2.10	1345	3	0.22	0.16
2.11	553	7	1.27	0.11
2.12	689	7	1.02	0.14
2.13	1377	7	0.51	0.10
2.14	1525	8	0.52	0.06
2.15	673	9	1.34	0.23
2.16	473	6	1.27	0.10
2.17	840	6	0.71	0.19
2.18	998	7	0.70	0.13
2.19	559	13	2.33	0.09
2.20	460	2	0.43	0.05
2.21	210	3	1.43	0.04

*1) These values originate from the count performed during the morphometric analysis
*2) These values originate from the additional analysis for finding cells within the initial cell size range only
RFI is calculated according to Eq. 1

1.5 R scripts

The R scripts used for analyzing the data are provided in PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.902769).

1.6 Detail on valve lengths

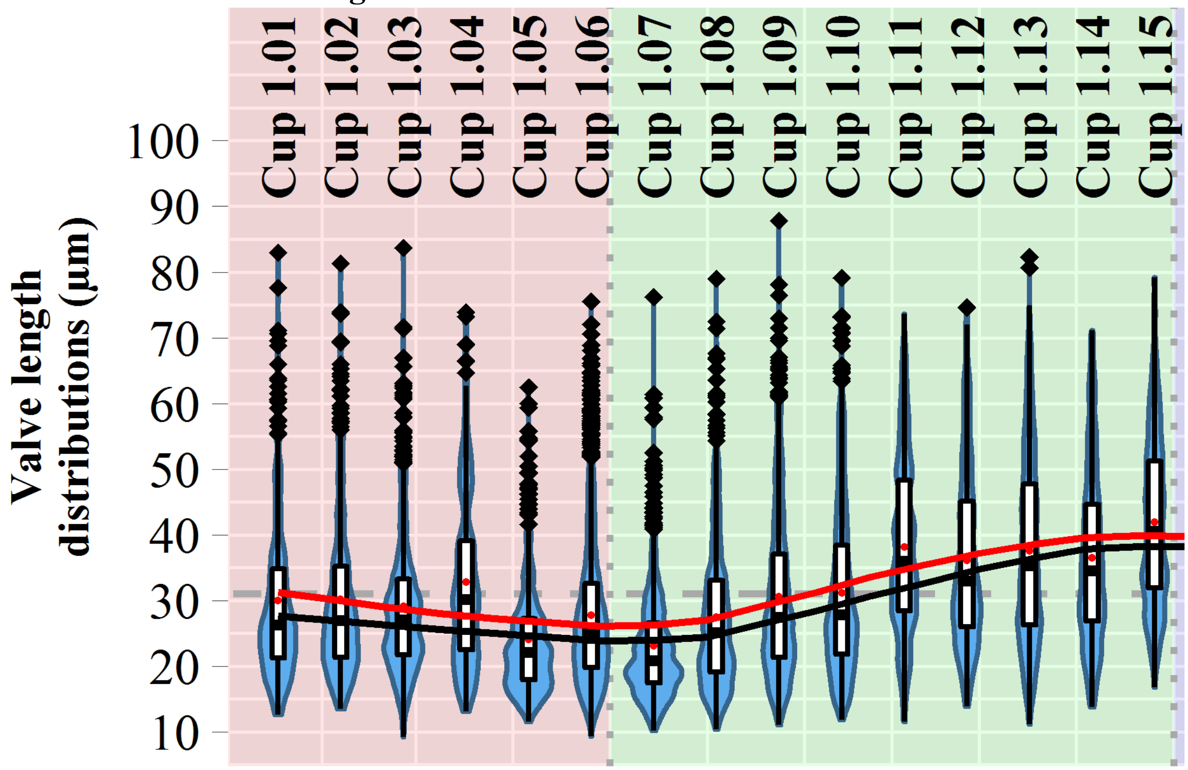

Fig. S1. Temporal variability of valve length distributions showing the ongoing loss of short valves during phase 2 of sampling year 2002/03 (detail from Fig. 4 c). Blue violin plots depict the valve length distributions, box plots represent quartiles, red dots indicate the mean, black dots inside the boxplots the median, red and black lines the respective trend lines derived by LOESS smoothing. The dashed line depicts 31 µm, the size limit below which sexual reproduction can occur. Phase 1 is highlighted by red, phase 2 by greed background.

1.7 Additional measurements

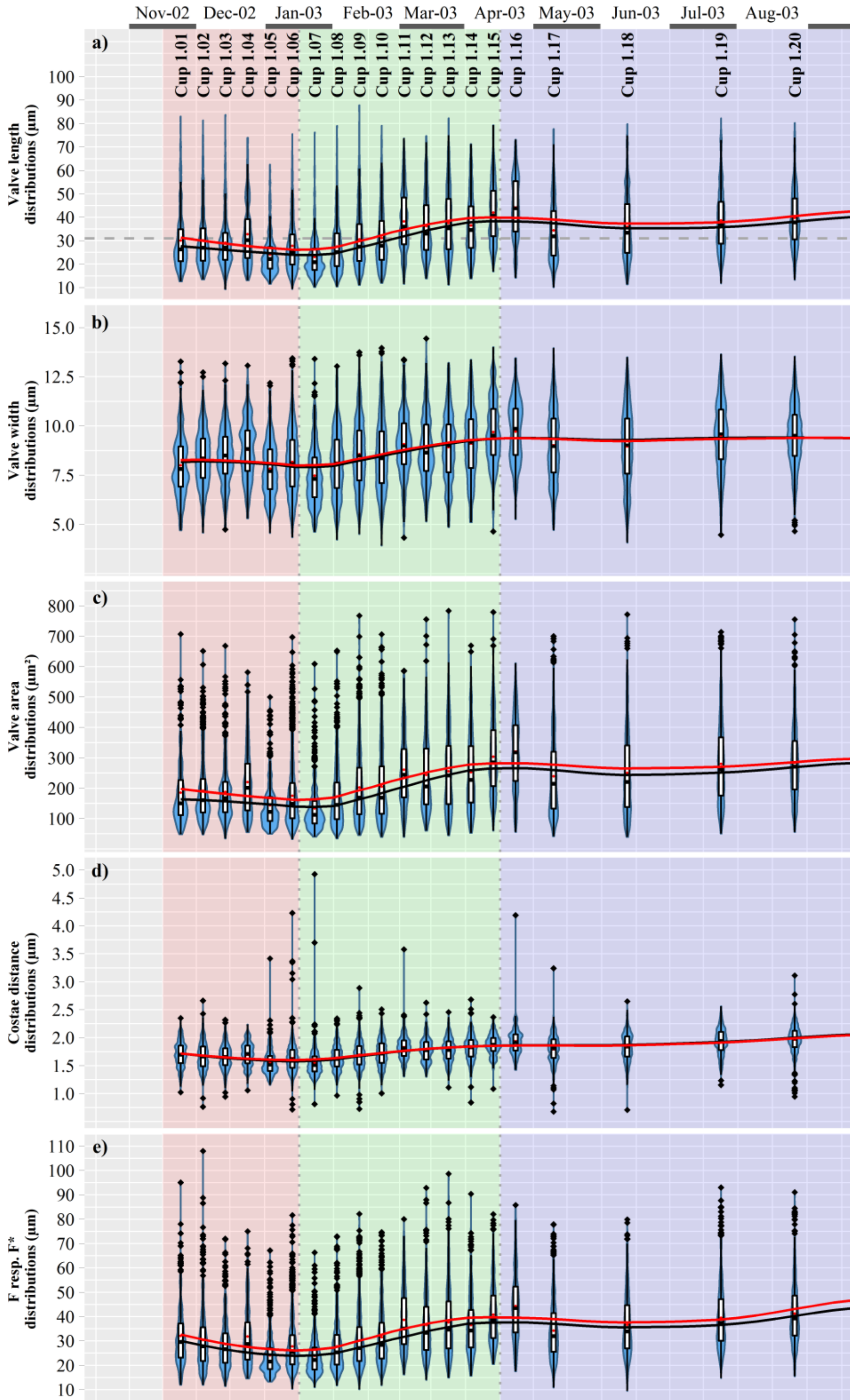

Fig. S2. Season 2002/03, distribution of a) valve length (cmp. Figs 4–5 c), b) valve width (transapical), c) valve area, d) costae distance, e) F (Fenner et al. 1976) respectively F* (Cortese & Gersonde 2007). To compensate for outliers, results for d) and e) with costae distances ≥ 5 µm were excluded.

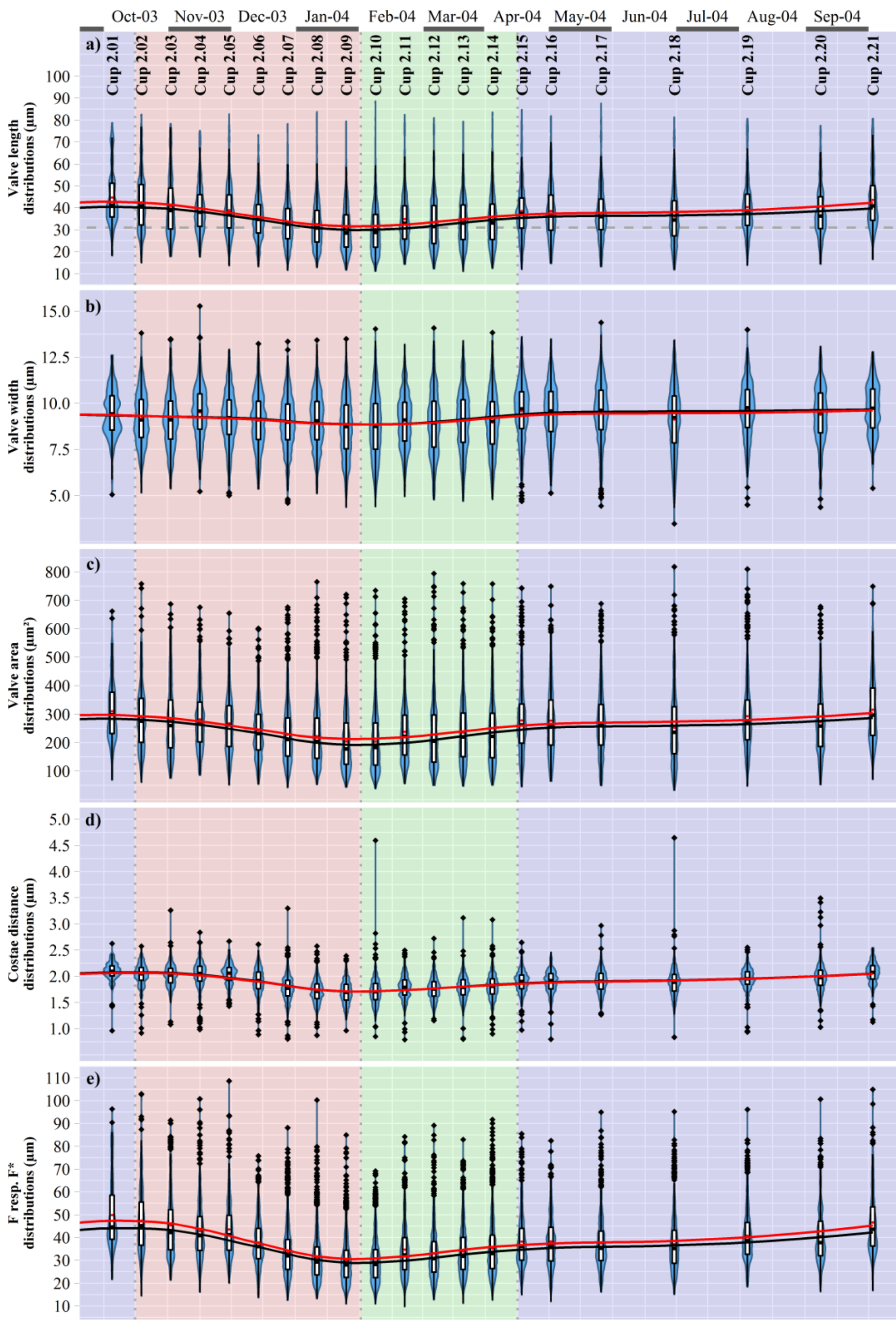

Fig. S3. Season 2003/04, distribution of a) valve length (cmp. Figs 4–5 c), b) valve width (transapical), c) valve area, d) costae distance, e) F (Fenner et al. 1976) respectively F* (Cortese & Gersonde 2007). To compensate for outliers, results for d) and e) with costae distances ≥ 5 µm were excluded.

1.8 Chlorophyll-a data

Table S2. Chlorophyll-*a* concentration derived from NASA's Giovanni online data system. NA = not available.

Date	Chlorophyll-*a* concentration ($mg \cdot m^{-3}$)
2002-08-16	0.138
2002-09-16	0.142
2002-10-16	0.153
2002-11-16	0.18
2002-12-16	0.182
2003-01-16	0.164
2003-02-15	0.255
2003-03-16	0.149
2003-04-16	0.119
2003-05-16	0.144
2003-06-16	NA
2003-07-16	NA
2003-08-16	0.125
2003-09-16	0.141
2003-10-16	0.167
2003-11-16	0.223
2003-12-16	0.179
2004-01-16	0.175
2004-02-15	0.114
2004-03-16	0.118
2004-04-16	0.118
2004-05-16	0.101
2004-06-16	NA
2004-07-16	NA
2004-08-16	0.111
2004-09-16	0.145
2004-10-16	0.162

1.9 Summary statistics of F. kerguelensis valve size distributions.

Table S3. Statistics of valve length measurements (compare Figs 4–5).

Cup	Middle day of sampling period	Length of sampling period (d)	Sample size	Valve length					% within sexually inducible size range [*1]	Rel. frequency of valves within the initial cell size range (RFI) [*2]
				Mean (µm)	Median (µm)	Skewness (µm)	Min (µm)	Max (µm)		
1.01	2002-11-23	10	300	30.0	26.3	1.4	13.0	70.7	66.7	0.06
1.02	2002-12-03	10	569	30.2	27.0	1.3	14.0	69.3	66.1	0.03
1.03	2002-12-13	10	525	29.1	27.0	1.4	13.1	63.0	68.0	0.04
1.04	2002-12-23	10	178	32.8	30.2	0.9	14.4	70.0	53.9	0.00
1.05	2003-01-02	10	455	24.1	22.1	1.6	12.7	54.6	85.1	0.07
1.06	2003-01-12	10	1630	27.7	25.3	1.2	11.9	61.6	70.9	0.02
1.07	2003-01-22	10	574	23.1	20.8	1.8	11.4	53.9	86.6	0.14
1.08	2003-02-01	10	736	27.5	25.1	1.2	12.0	62.8	69.4	0.02
1.09	2003-02-11	10	1010	30.5	27.4	1.0	12.8	65.6	60.6	0.04
1.10	2003-02-21	10	907	31.1	27.8	1.0	13.6	65.8	59.6	0.03
1.11	2003-03-03	10	338	38.1	35.9	0.4	14.7	67.8	33.4	0.00
1.12	2003-03-13	10	409	36.1	33.0	0.7	15.6	70.7	44.5	0.00
1.13	2003-03-23	10	478	37.5	35.3	0.5	14.7	70.7	39.5	0.02
1.14	2003-04-02	10	420	36.4	34.5	0.6	15.0	68.1	40.5	0.03
1.15	2003-04-12	10	379	41.9	40.6	0.4	20.3	74.0	23.0	0.05
1.16	2003-04-22	10	145	44.0	43.9	-0.1	17.5	70.0	17.9	0.00
1.17	2003-05-09	25	793	34.3	31.8	0.7	13.8	71.5	47.7	0.12
1.18	2003-06-11	40	592	35.9	33.3	0.6	13.1	72.2	44.8	0.18
1.19	2003-07-23	45	955	38.6	36.8	0.6	16.6	73.8	33.0	0.15
1.20	2003-08-25	20	811	39.7	37.8	0.6	17.1	74.0	26.4	0.28
2.01	2003-10-04	14	160	44.7	41.1	0.8	21.9	75.3	7.5	0.14
2.02	2003-10-18	14	327	42.4	40.9	0.6	20.1	76.0	21.4	0.18
2.03	2003-11-01	14	430	40.8	38.7	0.7	18.7	75.4	29.3	0.04
2.04	2003-11-15	14	600	39.6	38.0	0.8	20.5	72.5	23.0	0.01
2.05	2003-11-29	14	467	38.9	37.6	0.8	19.3	74.2	25.7	0.09
2.06	2003-12-13	14	718	35.6	34.7	0.6	16.6	62.3	34.0	0.03
2.07	2003-12-27	14	1639	33.6	32.5	0.8	15.8	66.2	44.9	0.40
2.08	2004-01-10	14	1360	32.5	30.8	1.1	15.5	69.6	50.9	0.19

2.09	2004-01-24	14	1228	30.5	28.5	1.0	13.6	66.5	58.7	0.18
2.10	2004-02-07	14	1345	30.3	28.8	0.9	13.3	62.5	57.3	0.16
2.11	2004-02-21	14	553	34.9	32.9	1.1	15.7	76.5	43.0	0.11
2.12	2004-03-06	14	689	33.3	31.5	1.0	13.7	74.9	48.2	0.14
2.13	2004-03-20	14	1377	34.1	33.0	0.8	15.1	69.3	44.0	0.10
2.14	2004-04-03	14	1525	34.7	33.1	0.8	14.7	72.1	42.8	0.06
2.15	2004-04-17	14	673	38.4	37.4	0.9	16.1	76.5	26.2	0.23
2.16	2004-05-01	14	473	38.5	36.9	0.7	15.7	76.3	29.6	0.10
2.17	2004-05-25	35	840	37.4	36.0	0.7	14.7	73.7	28.8	0.19
2.18	2004-06-29	35	998	36.0	34.4	0.8	13.8	74.1	37.0	0.13
2.19	2004-08-03	35	559	40.3	38.5	1.0	18.7	77.2	20.8	0.09
2.20	2004-09-07	35	460	38.4	36.3	0.8	16.5	73.6	27.8	0.05
2.21	2004-10-02	14	210	43.5	40.9	0.8	23.0	76.7	14.8	0.04

*1) refers to valves < 31.5 µm apical length
*2) see Equation 1

For cup 1.04, sampling size was low (n=178), even though this sample was taken within the productive period in late December 2002. Compared to the adjacent samples, valve density was unexpectedly sparse on the according slide. Also for cups 1.16, 2.01 and 2.21 sampling size was low (< 300), but here the slide density approximately accorded with the adjacent samples. We nevertheless decided to keep these measurements, since for the discussion of our results we mostly relate to smoothed general trends, which are not significantly changed by these samples.

1.10 Flux data

Table S4. Flux data from (Rigual-Hernández et al. 2015). Acronyms: $BSiO_2$, biogenic silica; POC, particulate organic carbon.

Cup	**Middle day of sampling period**	**Length of sampling period (d)**	**Total Mass Flux (mg · m^{-2} · d^{-1})**	**$BSiO_2$ (mg · m^{-2} · d^{-1})**	**$BSiO_2$ (%)**	**$CaCO_3$ (mg · m^{-2} · d^{-1})**	**$CaCO_3$ (%)**	**POC (mg · m^{-2} · d^{-1})**	**POC (%)**	***F. kerguelsnsis* flux (10^6 valves · m^{-2} · d^{-1})**	***F. kerguelensis* percentage of all diatoms**
1.01	2002-11-23	10.0	96.7	46.8	48	18.3	19	3.6	4	3.2	56.0
1.02	2002-12-03	10.0	190.6	110.6	58	37.5	20	4.6	2	12.1	43.6
1.03	2002-12-13	10.0	363.9	203.6	56	81.0	22	12.3	3	25.1	38.4
1.04	2002-12-23	10.0	240.7	134.0	56	47.2	20	11.4	5	9.5	26.3
1.05	2003-01-02	10.0	125.6	54.4	43	45.4	36	4.8	4	3.0	44.3
1.06	2003-01-12	10.0	158.9	93.7	59	35.0	22	5.4	3	13.2	71.7
1.07	2003-01-22	10.0	360.2	219.1	61	68.9	19	8.6	2	21.1	73.3
1.08	2003-02-01	10.0	213.8	113.7	53	48.3	23	5.5	3	31.4	69.7
1.09	2003-02-11	10.0	396.5	253.8	64	56.0	14	5.1	1	20.9	74.3

1.10	2003-02-21	10.0	60.8	26.6	44	23.7	39	2.2	4	6.9	40.5
1.11	2003-03-03	10.0	58.5	16.8	29	23.8	41	5.5	9	1.3	67.3
1.12	2003-03-13	10.0	89.3	26.6	30	37.0	41	7.0	8	0.9	63.1
1.13	2003-03-23	10.0	49.5	13.6	28	25.1	51	3.4	7	1.5	64.0
1.14	2003-04-02	10.0	33.5	7.6	23	18.2	54	2.7	8	0.8	78.0
1.15	2003-04-12	10.0	31.9	6.2	19	20.1	63	1.6	5	1.0	84.5
1.16	2003-04-22	10.0	21.6	5.8	27	14.0	65	1.0	4	0.6	78.7
1.17	2003-05-09	25.0	23.8	7.3	31	12.8	54	0.9	4	1.7	85.8
1.18	2003-06-11	40.0	23.7	10.8	46	7.8	33	1.0	4	1.4	86.7
1.19	2003-07-23	45.0	16.9	8.0	47	5.5	33	0.7	4	1.0	83.7
1.20	2003-08-25	20.0	33.6	13.5	40	11.9	35	1.8	5	2.1	81.0
2.01	2003-10-04	14.0	15.1	8.1	54	4.5	30	0.5	3	0.3	76.2
2.02	2003-10-18	14.0	23.4	12.6	54	7.0	30	0.8	3	0.8	73.9
2.03	2003-11-01	14.0	34.8	18.6	54	10.8	31	1.1	3	2.2	77.0
2.04	2003-11-15	14.0	121.0	54.6	45	19.9	16	2.0	2	9.9	85.8
2.05	2003-11-29	14.0	75.5	35.9	48	14.6	19	1.6	2	6.3	80.8
2.06	2003-12-13	14.0	178.6	111.6	62	40.2	23	3.6	2	10.9	55.0
2.07	2003-12-27	14.0	232.6	134.8	58	43.6	19	7.2	3	5.0	48.7
2.08	2004-01-10	14.0	182.0	114.5	63	33.7	19	6.2	3	20.2	65.2
2.09	2004-01-24	14.0	121.9	81.7	67	21.0	17	3.3	3	11.7	69.6
2.10	2004-02-07	14.0	63.9	37.0	58	10.8	17	2.7	4	7.3	65.7
2.11	2004-02-21	14.0	28.5	16.6	58	7.0	24	1.0	4	4.4	74.1
2.12	2004-03-06	14.0	31.0	18.4	59	7.7	25	0.9	3	4.6	73.7
2.13	2004-03-20	14.0	66.2	45.3	68	10.2	15	1.0	2	7.4	75.2
2.14	2004-04-03	14.0	84.6	52.6	62	16.9	20	0.8	1	8.9	76.9
2.15	2004-04-17	14.0	15.6	10.1	65	4.9	31	0.8	5	0.5	80.8
2.16	2004-05-01	14.0	10.6	6.9	65	4.0	38	0.5	4	0.6	87.2
2.17	2004-05-25	35.0	19.6	13.7	70	3.5	18	0.4	2	2.5	73.0
2.18	2004-06-29	35.0	19.3	11.3	59	4.5	23	1.1	6	1.2	75.8
2.19	2004-08-03	35.0	14.5	6.0	41	2.8	20	2.0	14	1.0	80.3
2.20	2004-09-07	35.0	10.3	5.0	49	3.8	37	0.5	5	1.0	90.3
2.21	2004-10-02	14.0	7.3	3.6	49	2.5	34	0.5	7	0.4	87.6

The flux data also is available at:
https://data.aad.gov.au/metadata/records/AAS_4078_diatoms_biogenic_flux_subantarctic/

1.11 KS test results regarding RI

The results of Kolmogorov-Smirnov tests, assessing the influence of adding results from subsequent ranking indices, can be found in the file "KS test results regarding RI.pdf".

1.12 References

BESZTERI B., ALLEN C., ALMANDOZ G.O., ARMAND L., BARCENA M.Á., CANTZLER H., CROSTA X., ESPER O., JORDAN R.W., KAUER G., KLAAS C., KLOSTER M., LEVENTER A., PIKE J. &

RIGUAL HERNÁNDEZ A.S. 2018. Quantitative comparison of taxa and taxon concepts in the diatom genus *Fragilariopsis*: A case study on using slide scanning, multi-expert image annotation and image analysis in taxonomy. *Journal of Phycology* 54: 703-719. doi: 10.1111/jpy.12767

BRADSKI G. & KAEHLER A. 2008. *Learning OpenCV : Computer vision with the OpenCV library.* O'Reilly, Sebastopol

BRESENHAM J.E. 1965. Algorithm for computer control of a digital plotter. *IBM Systems journal* 4: 25-30. doi: 10.1147/sj.41.0025

CORTESE G. & GERSONDE R. 2007. Morphometric variability in the diatom *Fragilariopsis kerguelensis*: Implications for Southern Ocean paleoceanography. *Earth and Planetary Science Letters* 257: 526-544. doi: 10.1016/j.epsl.2007.03.021

FENNER J., SCHRADER H. & WIENIGK H. 1976. Diatom phytoplankton studies in the southern Pacific Ocean, composition and correlation to the Antarctic Convergence and its paleoecological significance. *Initial Reports of the Deep Sea Drilling Project* 35: 757-813. Doi: 10.2973/dsdp.proc.35.app3.1976

KLOSTER M., ESPER O., KAUER G. & BESZTERI B. 2017. Large-scale permanent slide imaging and image analysis for diatom morphometrics. *Applied Sciences* 7: 330. doi: 10.3390/app7040330

RIGUAL-HERNÁNDEZ A.S., TRULL T.W., BRAY S.G., CORTINA A. & ARMAND L.K. 2015. Latitudinal and temporal distributions of diatom populations in the pelagic waters of the Subantarctic and Polar Frontal Zones of the Southern Ocean and their role in the biological pump. *Biogeosciences Discuss.* 12: 8615-8690. doi: 10.5194/bgd-12-8615-2015