Supplemental Information for:

Shotgun ancient DNA, pollen and macrofossil analysis of Lateglacial lake sediments from southern Sweden

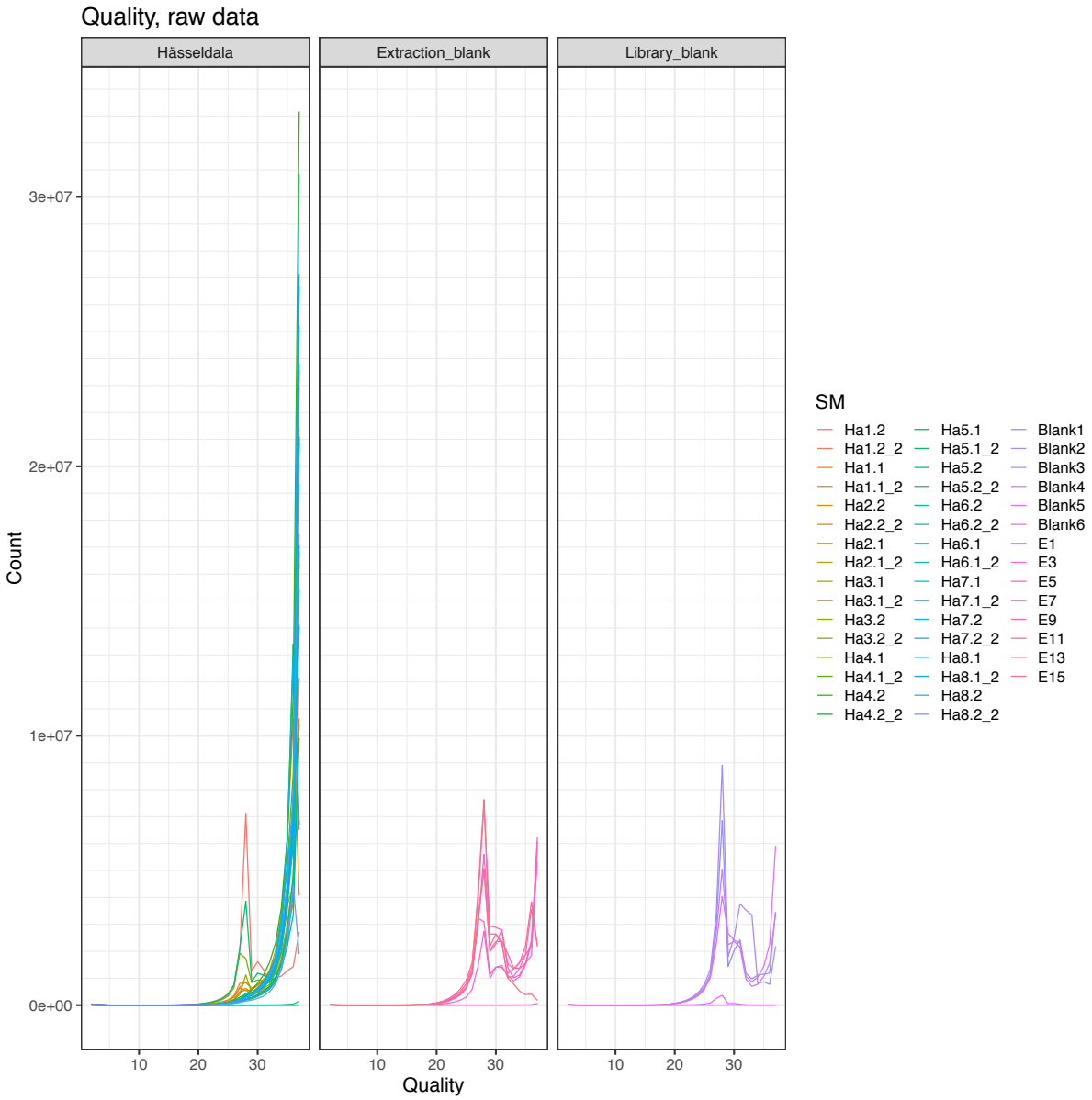
Laura Parducci, Mikkel W. Pedersen, Per Unneberg, Lu Han, Youri Lammers, Inger Greve Alsos, J. Sakari Salonen, Minna Väliranta, Tanja Slotte, Barbara Wohlfarth

Supplementary Figures

Figure S1

Raw (a) and trimmed (b) sequence data per sequence quality scores.

a



b

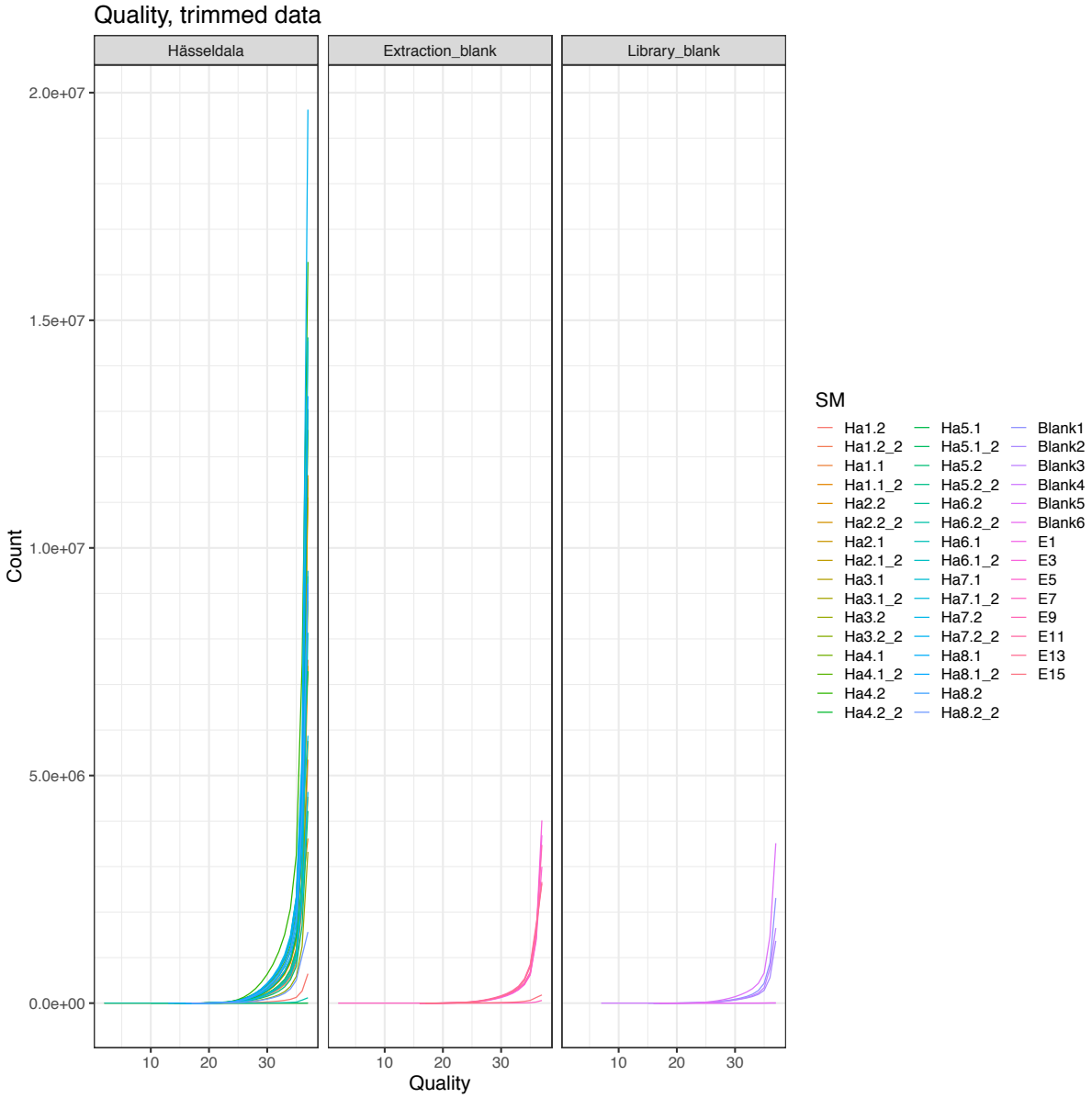
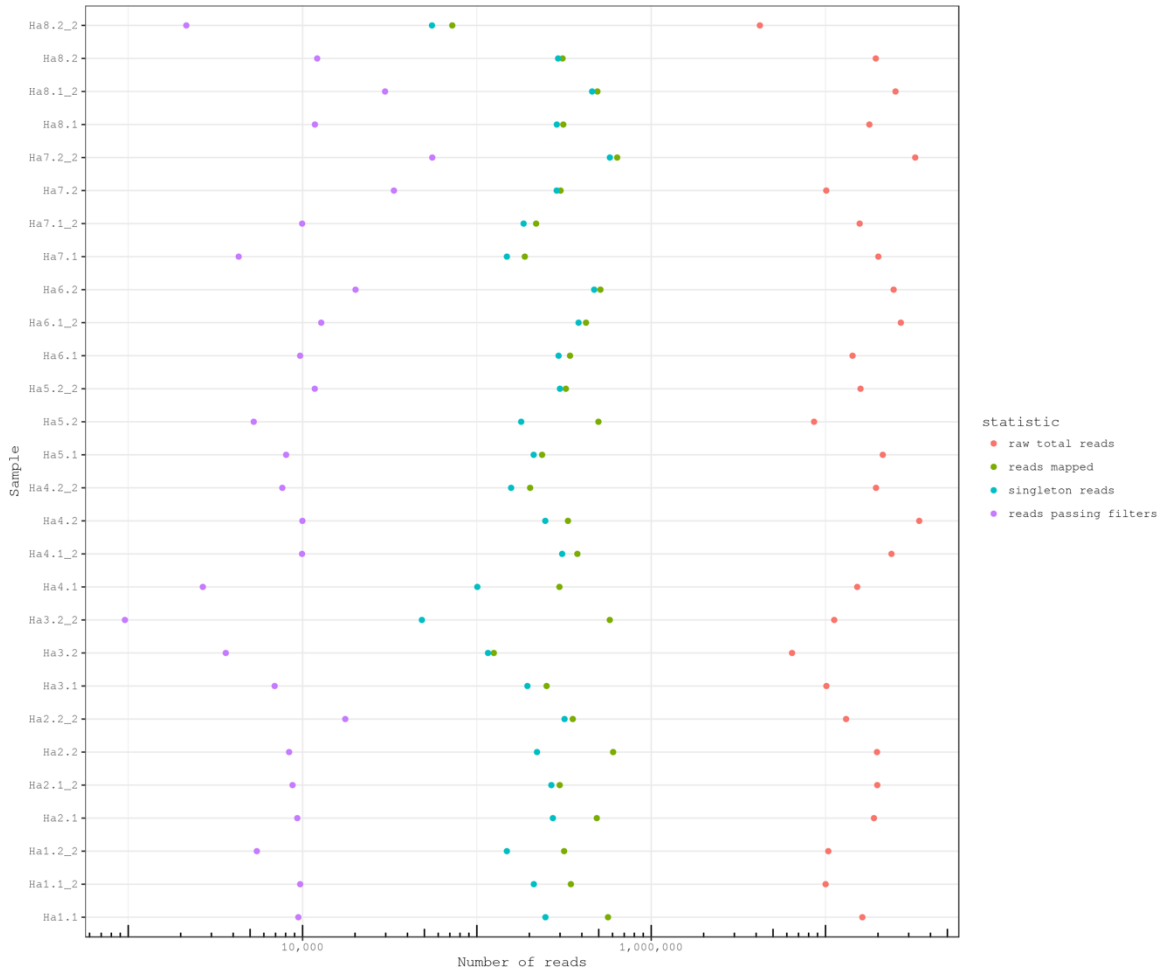

Quality, trimmed data
Hässeldala
Extraction_blank
Library_blank
Count
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
10
20
30
Quality
SM
Ha1.2
Ha1.2_2
Ha1.1
Ha1.1_2
Ha2.2
Ha2.2_2
Ha2.1
Ha2.1_2
Ha3.1
Ha3.1_2
Ha3.2
Ha3.2_2
Ha4.1
Ha4.1_2
Ha4.2
Ha4.2_2
Ha5.1
Ha5.1_2
Ha5.2
Ha5.2_2
Ha6.2
Ha6.2_2
Ha6.1
Ha6.1_2
Ha7.1
Ha7.1_2
Ha7.2
Ha7.2_2
Ha8.1
Ha8.1_2
Ha8.2
Ha8.2_2
Blank1
Blank2
Blank3
Blank4
Blank5
Blank6
E1
E3
E5
E7
E9
E11
E13
E15

Figure S2

Total number of reads, mapped reads, reads after duplication and reads after filtering (a) and read length distribution per sample of trimmed data as summarized by FastQC (b).

a



b

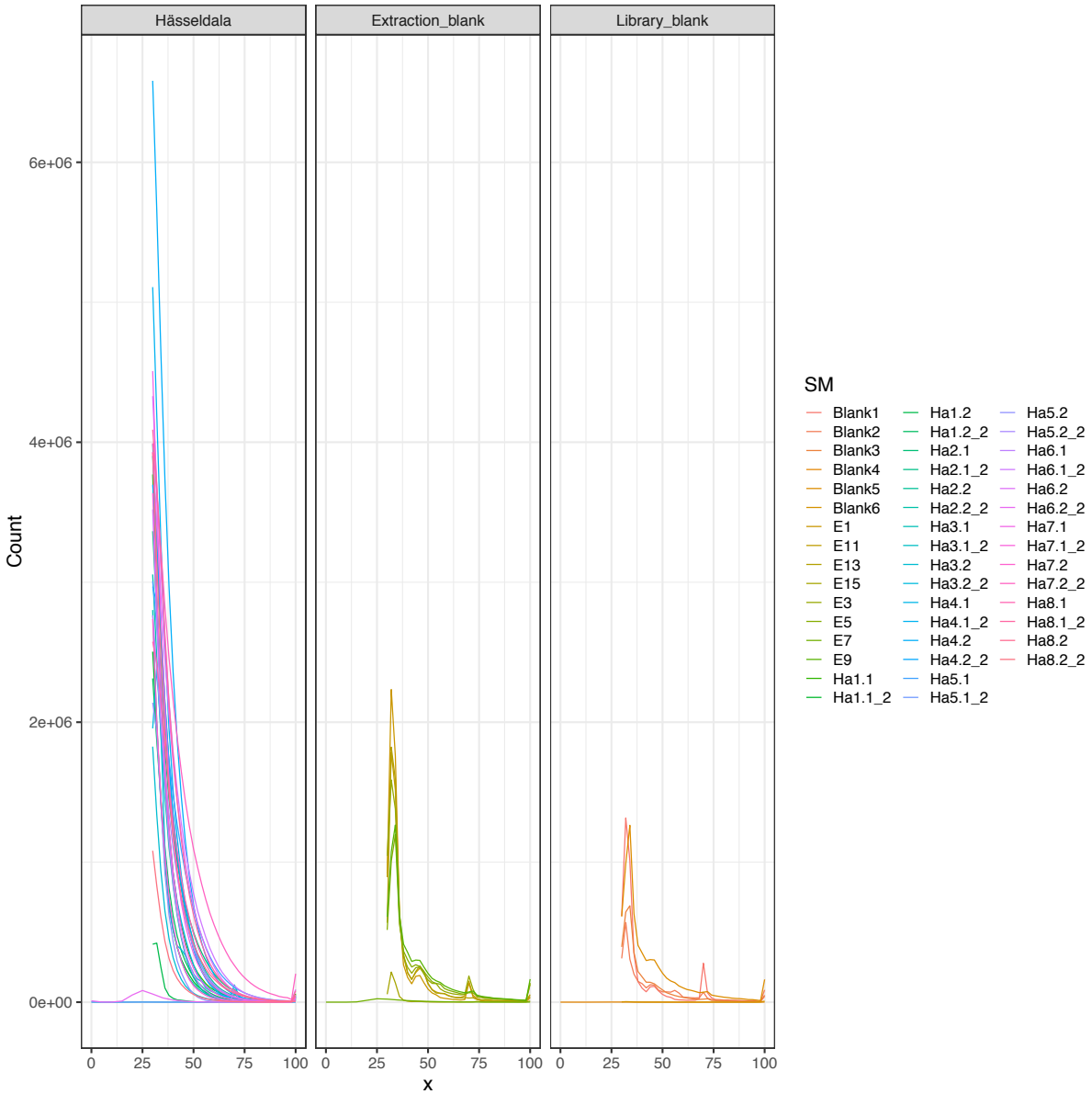
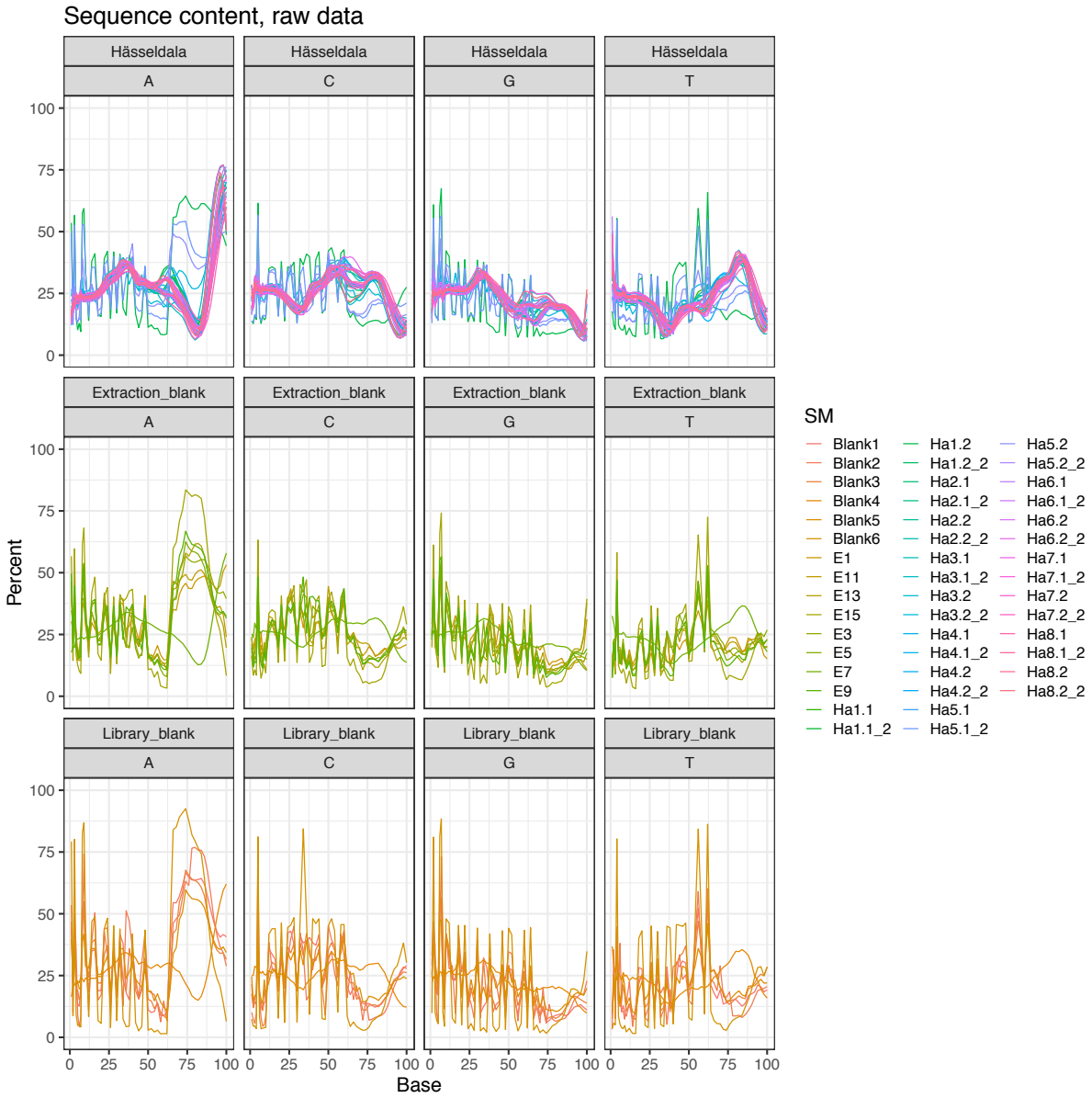

Hässeldala
Extraction_blank
Library_blank
Count
6e+06
4e+06
2e+06
0e+00
0
25
50
75
100
x
SM
Blank1
Blank2
Blank3
Blank4
Blank5
Blank6
E1
E11
E13
E15
E3
E5
E7
E9
Ha1.1
Ha1.1_2
Ha1.2
Ha1.2_2
Ha2.1
Ha2.1_2
Ha2.2
Ha2.2_2
Ha3.1
Ha3.1_2
Ha3.2
Ha3.2_2
Ha4.1
Ha4.1_2
Ha4.2
Ha4.2_2
Ha5.1
Ha5.1_2
Ha5.2
Ha5.2_2
Ha6.1
Ha6.1_2
Ha6.2
Ha6.2_2
Ha7.1
Ha7.1_2
Ha7.2
Ha7.2_2
Ha8.1
Ha8.1_2
Ha8.2
Ha8.2_2

Figure S3

Raw (a) and trimmed (b) sequence data per base sequence content. The Hässeldala samples display a sharp pronounced increase in thymines at the 5'-end and a longer increase in adenines towards the 3'-end. Intermediate positions show less variation than for raw reads.

a



b

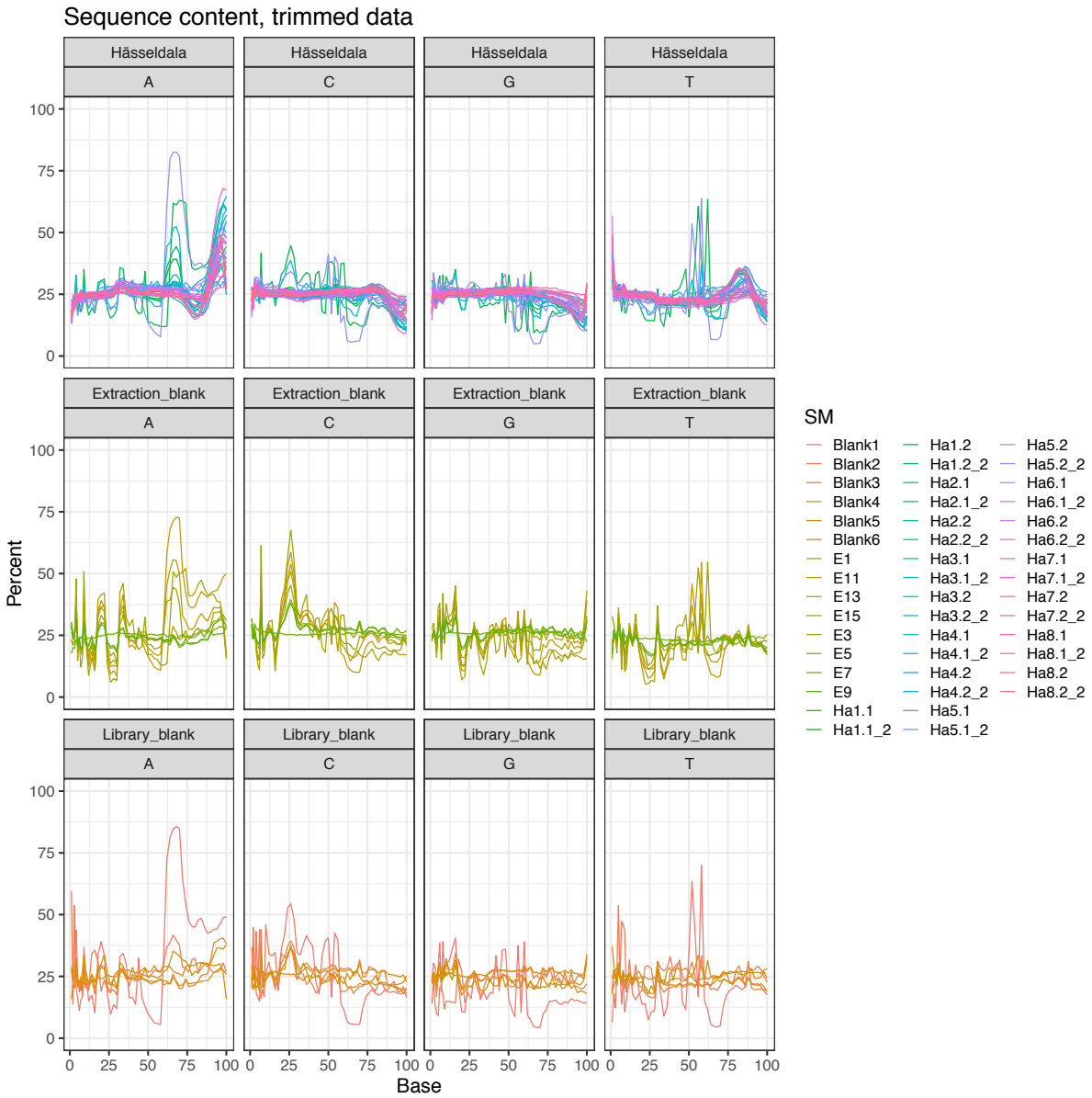

Sequence content, trimmed data
Hässeldala
A
C
G
T
Extraction_blank
Library_blank
Percent
Base
0
25
50
75
100
SM
Blank1
Blank2
Blank3
Blank4
Blank5
Blank6
E1
E11
E13
E15
E3
E5
E7
E9
Ha1.1
Ha1.1_2
Ha1.2
Ha1.2_2
Ha2.1
Ha2.1_2
Ha2.2
Ha2.2_2
Ha3.1
Ha3.1_2
Ha3.2
Ha3.2_2
Ha4.1
Ha4.1_2
Ha4.2
Ha4.2_2
Ha5.1
Ha5.1_2
Ha5.2
Ha5.2_2
Ha6.1
Ha6.1_2
Ha6.2
Ha6.2_2
Ha7.1
Ha7.1_2
Ha7.2
Ha7.2_2
Ha8.1
Ha8.1_2
Ha8.2
Ha8.2_2

Figure S4

K-mer analysis based on samples with at least 1 million raw reads. PCA of the distance matrix for k=6, raw data. PC1 mostly separates library and extraction blanks from Hässeldala samples. The k-mer profiles of sample HÄ1.2 resemble blanks indicating similar sequence content (it failed eventually to provide sequences), possibly due to contamination or that only a few k-mers, such as primer-dimers or adapter sequence, account for the separation. A similar case, albeit less clear, is for samples HÄ3.2_2 and HÄ5.2, which showed eventually to include a high number of low-quality bases.

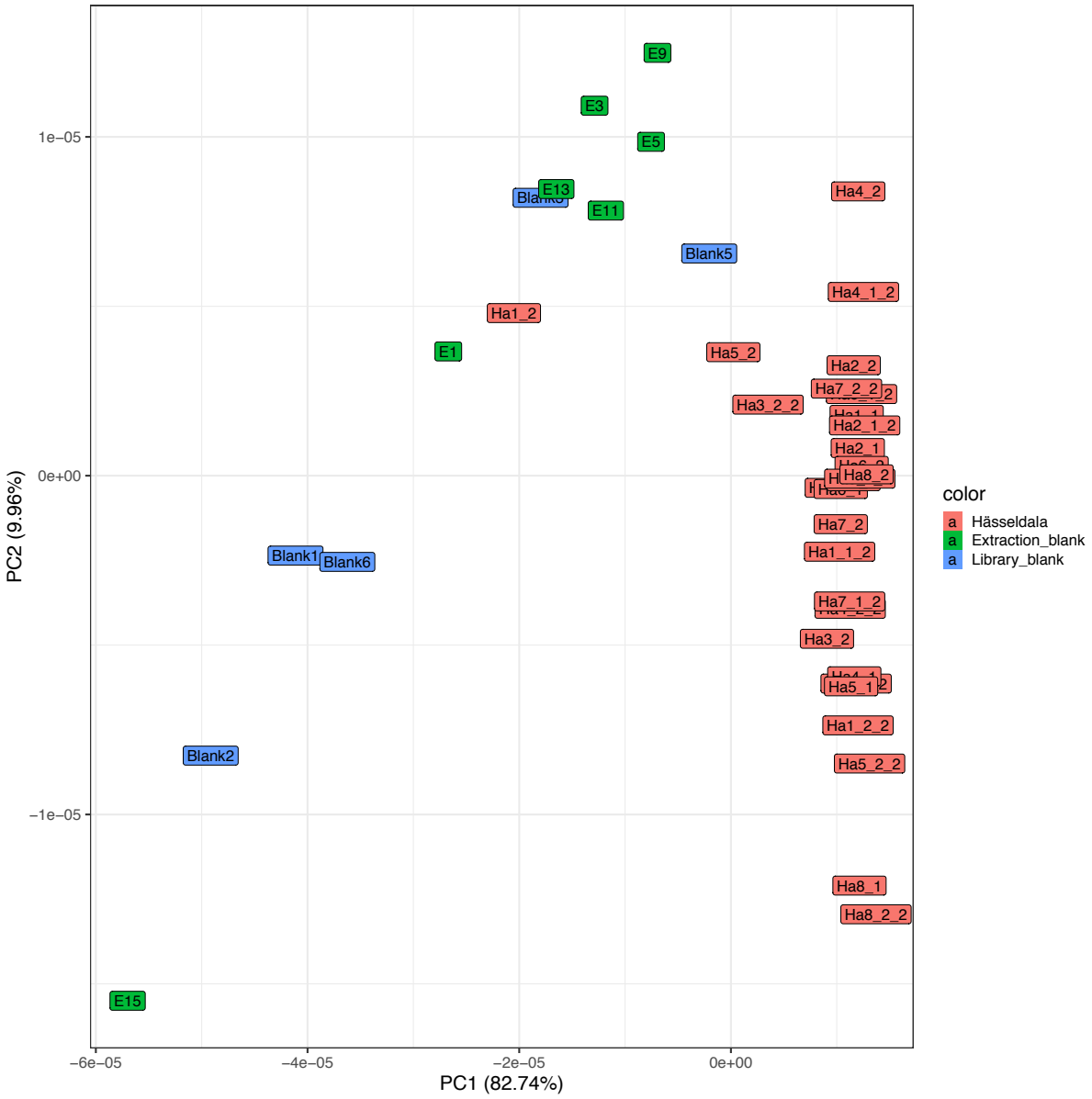


Table S1 (see separate excel file)
Hässeldala cores #7.4 and #8 all data DNA excluded

Table S2
Total number of raw reads and reads passing quality control filters. HÄ1.2 and HÄ6.2_2 failed to produce sufficient number of high-quality filtered sequences. Blank4, Hä3.1_2, Hä5.1_2 and E7 were removed due to low raw sequence content

	Sample	Total number of reads	Trimmed reads	%	Fail
1	Blank1	27590390	4681867	17	
2	Blank2	35291419	2987470	8.5	
3	Blank3	26002977	3620613	13.9	
4	Blank4	3571	3571	100	*
5	Blank5	29049772	7292930	25.1	
6	Blank6	1038889	25679	2.5	
7	E1	22253774	7296130	32.8	
8	E11	25549746	7204324	28.2	
9	E13	30382423	6865651	22.6	
10	E15	25852364	530053	2.1	
11	E3	33005067	7649384	23.2	
12	E5	31793332	6559444	20.6	
13	E7	112959	112959	100	*
14	E9	37762070	7348162	19.5	
15	Ha1.1	39619993	16267298	41.1	
16	Ha1.1_2	31869703	10012324	31.4	
17	Ha1.2	23703421	1375490	5.8	*
18	Ha1.2_2	25151161	10377364	41.3	
19	Ha2.1	39536594	18955470	47.9	
20	Ha2.1_2	41764765	19810677	47.4	
21	Ha2.2	42747103	19737050	46.2	
22	Ha2.2_2	29606647	13120745	44.3	
23	Ha3.1	35312295	10118363	28.7	
24	Ha3.1_2	5329	5329	100	*
25	Ha3.2	26741715	6427575	24	
26	Ha3.2_2	29076788	11217694	38.6	
27	Ha4.1	29310929	15199966	51.9	
28	Ha4.1_2	49987577	23898542	47.8	
29	Ha4.2	62160291	34478921	55.5	
30	Ha4.2_2	35295613	19490050	55.2	
31	Ha5.1	33733393	21303605	63.2	

32	Ha5.1_2	3326	3326	100	*
33	Ha5.2	29954087	8595682	28.7	
34	Ha5.2_2	28949491	15881157	54.9	
35	Ha6.1	36273097	14290138	39.4	
36	Ha6.1_2	46237277	27045716	58.5	
37	Ha6.2	41704341	24591978	59	
38	Ha6.2_2	275155	275155	100	*
39	Ha7.1	40014647	20091727	50.2	
40	Ha7.1_2	34572455	15700085	45.4	
41	Ha7.2	34997961	10099633	28.9	
42	Ha7.2_2	49222348	32701281	66.4	
43	Ha8.1	26380308	17850301	67.7	
44	Ha8.1_2	41749294	25208179	60.4	
45	Ha8.2	40278622	19427686	48.2	
46	Ha8.2_2	12668453	4202411	33.2	
47	Total	1364592932	549939155	40.3	
48	Total (samples)	1038904179	487760918	46.9	
49	Total (controls)	325688753	62178237	19.1	

Table S3
Mapping and filtering summary for the Hässeldala samples.

Sample	QC reads	Mapped reads	%	Singleton Reads	Reads passing filters	%
Ha1.1	16267298	565279	0.035	247372	9460	0.017
Ha1.1_2	10012324	346125	0.035	212220	9690	0.028
Ha1.2_2	10377364	316475	0.030	148636	5466	0.017
Ha2.1	18955470	487455	0.026	272961	9328	0.019
Ha2.1_2	19810677	298669	0.015	267487	8761	0.029
Ha2.2	19737050	604723	0.031	221259	8374	0.014
Ha2.2_2	13120745	354852	0.027	318245	17583	0.050
Ha3.1	10118363	251239	0.025	194932	6917	0.028
Ha3.2	6427575	125077	0.019	115818	3627	0.029
Ha3.2_2	11217694	578538	0.052	48368	957	0.002
Ha4.1	15199966	297587	0.020	100600	2678	0.009
Ha4.1_2	23898542	377154	0.016	308528	9933	0.026
Ha4.2	34478921	333001	0.010	246872	9979	0.030
Ha4.2_2	19490050	202038	0.010	157210	7655	0.038
Ha5.1	21303605	236650	0.011	211786	8049	0.034
Ha5.2	8595682	498161	0.058	179547	5247	0.011
Ha5.2_2	15881157	323908	0.020	299541	11750	0.036
Ha6.1	14290138	342269	0.024	294267	9685	0.028
Ha6.1_2	27045716	422677	0.016	383131	12798	0.030
Ha6.2	24591978	511352	0.021	470894	20135	0.039
Ha7.1	20091727	188342	0.009	148645	4301	0.023
Ha7.1_2	15700085	218698	0.014	185388	9945	0.045
Ha7.2	10099633	301967	0.030	286896	33429	0.111
Ha7.2_2	32701281	637985	0.020	579271	55470	0.087
Ha8.1	17850301	313135	0.018	287491	11774	0.038
Ha8.1_2	25208179	491370	0.019	458193	29737	0.061
Ha8.2	19427686	310144	0.016	292432	12135	0.039
Ha8.2_2	4202411	72249	0.017	55273	2154	0.030
Tot/average	**486101618**	**10007119**	**0.023**	**6993263**	**337017**	**0.034**

Table S4 (see separate excel file)
Custom reference dataset containing the plastid sequences from the NCBI RefSeq project (2477 sequences, 29th January 2018) and the plastid and ribosomal sequences from the PhyloNorway project (259 plant species from northern Norway included).

Table S5
Mapping summary based on mapped reads passing filters and taxonomic assigmnents. The taxonomy columns correspond to the number of mapped reads (and percentages) that have been assigend to a taxonomy class.

	Mapped reads passing filter	Mapped reads passing filter (%)	Archaea	Archaea (%)	Bacteria	Bacteria (%)	Eukaryote	Eukaryote (%)
Blank1	110531	100	45	0.04	74690	67.57	35464	32.09
Blank2	110925	100	18	0.02	102629	92.52	8205	7.40
Blank3	323424	100	329	0.10	265568	82.11	56727	17.54
Blank5	712096	100	221	0.03	603506	84.75	107097	15.04
Blank6	2223	100	7	0.31	864	38.87	1350	60.73
E1	119770	100	64	0.05	83349	69.59	35973	30.04
E11	978694	100	63	0.01	281725	28.79	696269	71.14
E13	928222	100	125	0.01	269327	29.02	658239	70.91
E15	19576	100	4	0.02	4747	24.25	14805	75.63
E3	520613	100	145	0.03	432719	83.12	75340	14.47
E5	699185	100	307	0.04	560979	80.23	136594	19.54
E9	731123	100	681	0.09	612597	83.79	116454	15.93
Ha1.1	120790	100	2849	2.36	112357	93.02	5262	4.36
Ha1.1_2	55828	100	2160	3.87	46947	84.09	6491	11.63
Ha1.2_2	30838	100	1565	5.07	25338	82.16	3811	12.36
Ha2.1	44539	100	3953	8.88	37401	83.97	2925	6.57
Ha2.1_2	10703	100	1603	14.98	8597	80.32	418	3.91
Ha2.2	101732	100	3244	3.19	90409	88.87	7899	7.76
Ha2.2_2	20600	100	1862	9.04	18237	88.53	417	2.02
Ha3.1	12572	100	2619	20.83	9136	72.67	707	5.62
Ha3.2	4391	100	507	11.55	3585	81.64	279	6.35
Ha3.2_2	206975	100	1170	0.57	194858	94.15	10921	5.28
Ha4.1	78708	100	1584	2.01	74913	95.18	2120	2.69
Ha4.1_2	16787	100	2600	15.49	13253	78.95	819	4.88

Ha4.2	15147	100	1269	8.38	13129	86.68	686	4.53
Ha4.2_2	12039	100	717	5.96	8648	71.83	2642	21.95
Ha5.1	9752	100	1930	19.79	7292	74.77	447	4.58
Ha5.2	23064	100	915	3.97	18250	79.13	3769	16.34
Ha5.2_2	13945	100	1747	12.53	11692	83.84	418	3.00
Ha6.1	13047	100	1341	10.28	10854	83.19	738	5.66
Ha6.1_2	16139	100	2330	14.44	12824	79.46	815	5.05
Ha6.2	23398	100	2692	11.51	19944	85.24	649	2.77
Ha7.1	5540	100	783	14.13	4531	81.79	186	3.36
Ha7.1_2	11358	100	632	5.56	9716	85.54	863	7.60
Ha7.2	35622	100	504	1.41	34846	97.82	229	0.64
Ha7.2_2	62541	100	982	1.57	60901	97.38	553	0.88
Ha8.1	14555	100	587	4.03	13545	93.06	359	2.47
Ha8.1_2	32359	100	917	2.83	30973	95.72	350	1.08
Ha8.2	13536	100	810	5.98	12244	90.46	422	3.12
Ha8.2_2	5518	100	184	3.33	4245	76.93	1071	19.41
Average				**5.61**		**78.77**		**15.16**
Total	6268405		46065		4201365		1998783	

Table S6 (see separate excel file)
Selected DNA reads mapping against the custom reference dataset containing the plastid sequences from the NCBI RefSeq project (2477 sequences, 29th January 2018) and the plastid and ribosomal sequences from the PhyloNorway project (259 plant species from Norway included).