W+ ARTIKEL
WISSENSCHAFTPLUS
LK-Verlags UG

Ein elementarer Ansatz am Beispiel von SARS-CoV-2

Strukturelle Analyse von Sequenzdaten in der Virologie

Tabellen und Abbildungen

Strukturelle Analyse von Sequenzdaten in der Virologie

Ein elementarer Ansatz am Beispiel von SARS-CoV-2

Tabellen und Abbildungen

von einem Mathematiker aus Hamburg,
der noch unbekannt bleiben möchte

Contig ID	Contig length	Abundance	GenBank accession	Subect title	Identity (%)	Length	E-value	Bitscore
k141_11989	29.802	120.341	MN908947.3	Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome	100,000	29.801	0,000E+00	55.033
k141_11989	29.802	120.341	OK372407.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/TN-VUMC-000018/2020, complete genome	100,000	29.801	0,000E+00	55.033
k141_11989	29.802	120.341	MG772933.1	Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome	89,120	28.464	0,000E+00	26.943
k141_11989	29.802	120.341	AY274119.3	SARS coronavirus Tor2, complete genome	82,300	26.631	0,000E+00	15.175
k141_11989	29.802	120.341	AY278741.1	SARS coronavirus Urbani, complete genome	82,300	26.632	0,000E+00	15.169
k141_5255	5.659	290.188	CP040006.1	Schaalia odontolytica strain XH001 chromosome, complete genome	96,364	5.115	0,000E+00	8.379
k141_12253	5.414	213.744	CP012410.1	Leptotrichia sp, oral taxon 212 strain W10393, complete genome	96,229	4.932	0,000E+00	8.059
k141_14387	3.526	933.303	CP003667.1	Prevotella sp, oral taxon 299 str, F0039 plasmid, complete sequence	93,381	2.100	0,000E+00	3.083
k141_14692	3.364	313.495	CP040504.1	Neisseria sp, oral taxon 014 str, F0314 chromosome	93,322	2.965	0,000E+00	4.357
k141_631	3.014	2.200.770	LR778174.1	Veillonella parvula strain SKV38 genome assembly, chromosome: 1	93,143	2.844	0,000E+00	4.150
k141_11680	2.964	116.125	AP019846.1	Leptotrichia hongkongensis JMUB5056 DNA, complete genome	97,569	2.962	0,000E+00	5.066
k141_27232	2.811	1.407.705	MF164264.1	Homo sapiens clone BAC JH4 genomic sequence	98,851	2.436	0,000E+00	4.337
k141_27232	2.811	1.407.705	NR_146117.1	Homo sapiens RNA, 45S pre-ribosomal N4 (RNA45SN4), ribosomal RNA	98,850	2.435	0,000E+00	4.335
k141_27232	2.811	1.407.705	CP068263.2	Homo sapiens isolate CHM13 chromosome 15	98,850	2.434	0,000E+00	4.335
k141_27232	2.811	1.407.705	AL353644.3	Human DNA sequence from clone RP11-164K15 on chromosome 22, complete sequence	98,850	2.435	0,000E+00	4.335
k141_27232	2.811	1.407.705	NR_046235.3	Homo sapiens RNA, 45S pre-ribosomal N5 (RNA45SN5), ribosomal RNA	98,809	2.435	0,000E+00	4.331
k141_16443	2.626	693.461	CP020566.1	Veillonella atypica strain OK5, complete genome	96,758	2.190	0,000E+00	3.633
k141_10436	2.373	425.138	LR134384.1	Prevotella oris strain NCTC13071 genome assembly, chromosome: 1	93,473	2.390	0,000E+00	3.530
k141_6751	2.309	99.053	CP068256.2	Homo sapiens isolate CHM13 chromosome 22	99,908	2.165	0,000E+00	3.988
k141_6751	2.309	99.053	CP068257.2	Homo sapiens isolate CHM13 chromosome 21	99,591	2.199	0,000E+00	4.004
k141_6751	2.309	99.053	CP068263.2	Homo sapiens isolate CHM13 chromosome 15	99,500	2.199	0,000E+00	3.993
k141_6751	2.309	99.053	AC231275.2	Homo sapiens FOSMID clone ABC12-46987300E12 from chromosome unknown, complete sequence	99,500	2.199	0,000E+00	3.991
k141_6751	2.309	99.053	MT497387.1	Homo sapiens clone BAC JH13 genomic sequence	99,454	2.199	0,000E+00	3.986
k141_10208	1.812	1.400.352	FM996435.1	Uncultured bacterium partial 16S rRNA gene, clone 16sps19-1g02,p1ka	97,011	1.405	0,000E+00	2.359
k141_12515	1.744	184.440	KY962518.1	Homo sapiens external transcribed spacer 18S ribosomal RNA gene, internal transcribed spacer 1, 5,8S ribosomal RNA gene, internal transcribed spacer 2, 28S ribosomal RNA gene, and external transcribed spacer, complete sequence	99,323	1.626	0,000E+00	2.942
k141_12515	1.744	184.440	MF164269.1	Homo sapiens clone BAC JH1 genomic sequence	99,323	1.626	0,000E+00	2.942
k141_12515	1.744	184.440	NR_146144.1	Homo sapiens RNA, 45S pre-ribosomal N2 (RNA45SN2), ribosomal RNA	99,323	1.626	0,000E+00	2.942
k141_12515	1.744	184.440	NR_146148.1	Homo sapiens RNA, 28S ribosomal N2 (RNA28SN2), ribosomal RNA	99,323	1.626	0,000E+00	2.942
k141_12515	1.744	184.440	NR_145822.1	Homo sapiens RNA, 28S ribosomal N1 (RNA28SN1), ribosomal RNA	99,323	1.626	0,000E+00	2.942
k141_28311	1.506	248.289	JQ460207.1	Uncultured bacterium clone 070054_332 16S ribosomal RNA gene, partial sequence	98,213	1.287	0,000E+00	2.248
k141_26154	1.211	136.391	JQ454767.1	Uncultured bacterium clone 069096_294 16S ribosomal RNA gene, partial sequence	99,459	925	0,000E+00	1.679
k141_24818	1.195	583.961	JQ460141.1	Uncultured bacterium clone 070054_143 16S ribosomal RNA gene, partial sequence	92,105	1.178	0,000E+00	1.637
k141_22555	1.131	232.205	CP003667.1	Prevotella sp, oral taxon 299 str, F0039 plasmid, complete sequence	94,420	1.129	0,000E+00	1.735
k141_8265	1.128	183.428	MN849515.1	Homo sapiens isolate NJ44 haplogroup P1d1 mitochondrion, complete genome	99,466	1.124	0,000E+00	2.041
k141_8265	1.128	183.428	MN398712.1	Homo sapiens isolate HG3118 haplogroup P1d1 mitochondrion, complete genome	99,466	1.124	0,000E+00	2.041
k141_8265	1.128	183.428	MN706604.1	Homo sapiens isolate BachoKiro_BK_1653 mitochondrion, complete genome	99,466	1.124	0,000E+00	2.041
k141_8265	1.128	183.428	MF437277.1	Homo sapiens isolate 250 mitochondrion, complete genome	99,466	1.124	0,000E+00	2.041
k141_8265	1.128	183.428	MK491356.1	Homo sapiens isolate 2_Mu mitochondrion, complete genome	99,466	1.124	0,000E+00	2.041
k141_11940	1.097	508.437	CP072350.1	Prevotella melaninogenica strain F0301 chromosome 2, complete sequence	99,544	1.097	0,000E+00	1.999
k141_10110	1.043	85.091	CP072347.1	Prevotella melaninogenica strain F0516 chromosome 2, complete sequence	91,383	882	0,000E+00	1.206
k141_5437	1.007	1.295.606	CP072360.1	Prevotella melaninogenica strain F0091 chromosome 1, complete sequence	98,512	1.008	0,000E+00	1.777
k141_20271	992	188.951	CP023863.1	Prevotella jejuni strain CD3:33 chromosome I, complete sequence	97,414	580	0,000E+00	985

Tabelle 1: *Die 50 assemblierten Contigs mit der größten Fülle mit Megahit (v.1.2.9), Datenbankabfrage vom 05.12.2021*

Contig ID	Contig length	Abundance	GenBank accession	Subect title	Identity (%)	Length	E-value	Bitscore
k141_4074	917	174.529	CP068256.2	Homo sapiens isolate CHM13 chromosome 22	97,686	778	0,000E+00	1.338
k141_4074	917	174.529	CP068257.2	Homo sapiens isolate CHM13 chromosome 21	97,686	778	0,000E+00	1.338
k141_4074	917	174.529	CP068263.2	Homo sapiens isolate CHM13 chromosome 15	97,686	778	0,000E+00	1.338
k141_4074	917	174.529	CP086022.1	Porphyrobacter sp, GA68 chromosome, complete genome	97,452	785	0,000E+00	1.338
k141_4074	917	174.529	AP025035.1	Citrobacter freundii RTE-E5 DNA, complete genome: sequence1	97,452	785	0,000E+00	1.338
k141_5448	913	78.694	CP024724.1	Prevotella intermedia strain KCOM 2837 chromosome 2, complete sequence	91,730	919	0,000E+00	1.264
k141_24026	850	134.614	CP085934.1	Prevotella copri DSM 18205 strain FDAARGOS_1573 plasmid unnamed2, complete sequence	97,291	406	0,000E+00	688
k141_4059	812	132.074	CP072363.1	Prevotella jejuni strain F0697 chromosome 1, complete sequence	98,684	532	0,000E+00	942
k141_525	754	217.291	CP072333.1	Porphyromonas sp, oral taxon 275 strain W7780 chromosome, complete genome	98,802	501	0,000E+00	891
k141_8586	656	762.405	CP016205.1	Prevotella scopos JCM 17725 strain W2052 chromosome 2 genome	98,176	658	0,000E+00	1.147
k141_19969	639	157.783	MT242596.1	Homo sapiens isolate DH001 mitochondrion, complete genome	100,000	477	0,000E+00	881
k141_19969	639	157.783	OL521838.1	Homo sapiens haplogroup I3 mitochondrion, complete genome	100,000	477	0,000E+00	881
k141_19969	639	157.783	OK104093.1	Homo sapiens haplogroup H4a1a1a mitochondrion, complete genome	100,000	477	0,000E+00	881
k141_19969	639	157.783	OK266950.1	Homo sapiens haplogroup H3i mitochondrion, complete genome	100,000	477	0,000E+00	881
k141_19969	639	157.783	OK239657.1	Homo sapiens haplogroup H1b1d mitochondrion, complete genome	100,000	477	0,000E+00	881
k141_534	620	199.625	CP072361.1	Prevotella melaninogenica strain F0054 chromosome 1, complete sequence	98,387	620	0,000E+00	1.090
k141_7697	569	360.511	CP023863.1	Prevotella jejuni strain CD3:33 chromosome I, complete sequence	96,473	567	0,000E+00	935
k141_11094	532	62.843	CP072331.1	Prevotella veroralis strain F0319 chromosome 2, complete sequence	97,543	529	0,000E+00	904
k141_17668	508	202.095	CP085941.1	Prevotella melaninogenica strain FDAARGOS_1567 chromosome 2, complete sequence	99,018	509	0,000E+00	911
k141_11371	427	198.847	CP072360.1	Prevotella melaninogenica strain F0091 chromosome 1, complete sequence	99,532	427	0,000E+00	778
k141_9606	408	63.287	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	98,529	408	0,000E+00	721
k141_25754	384	133.274	MN297237.1	Homo sapiens LHRI_LNC32,2 lncRNA gene, complete sequence	100,000	256	4,700E-129	473
k141_25754	384	133.274	MN297236.1	Homo sapiens LHRI_LNC32,1 lncRNA gene, complete sequence	100,000	256	4,700E-129	473
k141_25754	384	133.274	CP034492.1	Eukaryotic synthetic construct chromosome 14	100,000	256	4,700E-129	473
k141_25754	384	133.274	NG_050638.2	Homo sapiens ribosomal protein S29 (RPS29), RefSeqGene (LRG_1147) on chromosome 14	100,000	256	4,700E-129	473
k141_25754	384	133.274	XR_001750762.1	PREDICTED: Homo sapiens uncharacterized LOC107987206 (LOC107987206), ncRNA	100,000	256	4,700E-129	473
k141_13347	351	163.140	JQ470050.1	Uncultured bacterium clone 071024_066 16S ribosomal RNA gene, partial sequence	100,000	351	0,000E+00	649
k141_17635	335	187.903	LT677940.1	Prevotella melaninogenica partial 16S rRNA gene, isolate 219N_3354	100,000	335	4,900E-173	619
k141_14693	329	151.450	CP023863.1	Prevotella jejuni strain CD3:33 chromosome I, complete sequence	98,160	326	4,910E-158	569
k141_21608	324	129.790	MW717453.1	Veillonella sp, strain KCOM 3945 16S ribosomal RNA gene, partial sequence	100,000	324	6,150E-167	599
k141_1252	321	182.075	CP023863.1	Prevotella jejuni strain CD3:33 chromosome I, complete sequence	99,688	321	1,320E-163	588
k141_10440	318	83.171	CP003667.1	Prevotella sp, oral taxon 299 str, F0039 plasmid, complete sequence	99,686	318	6,070E-162	582
k141_14695	318	86.326	CP023863.1	Prevotella jejuni strain CD3:33 chromosome I, complete sequence	99,371	318	2,820E-160	577
k141_10440	318	83.171	AP024484.1	Prevotella sp, WR041 DNA, complete genome	94,688	320	2,930E-135	494
k141_2510	286	83128	CP023864.1	Prevotella jejuni strain CD3:33 chromosome II, complete sequence	100,000	286	7,110E-146	529
k141_19087	281	62524	LR778174.1	Veillonella parvula strain SKV38 genome assembly, chromosome: 1	100,000	281	4,200E-143	520
k141_6152	245	186296	CP023864.1	Prevotella jejuni strain CD3:33 chromosome II, complete sequence	99,190	247	2,220E-120	444
k141_5974	234	121016	LC358497.1	Uncultured bacterium 62MG04014 gene for 16S rRNA, partial sequence	100,000	234	4,550E-117	433
k141_17500	232	117574	MF801036.1	Uncultured bacterium clone saliva72 16S ribosomal RNA gene, partial sequence	100,000	232	5,820E-116	429
k141_17824	208	63938	CP072331.1	Prevotella veroralis strain F0319 chromosome 2, complete sequence	98,387	186	1,930E-85	327
k141_7700	206	112696	CP085941.1	Prevotella melaninogenica strain FDAARGOS_1567 chromosome 2, complete sequence	100,000	206	1,440E-101	381

Contig ID	Contig length	GenBank accession	Subect title	Identity (%)	Length	E-value	Bitscore
k141_11989	29.802	MN908947.3	Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome	100,000	29.801	0,000E+00	55.033
k141_11989	29.802	OK372407.1	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/TN-VUMC-000018/2020, complete genome	100,000	29.801	0,000E+00	55.033
k141_11989	29.802	MG772933.1	Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome	89,120	28.464	0,000E+00	26.943
k141_11989	29.802	AY274119.3	SARS coronavirus Tor2, complete genome	82,300	26.631	0,000E+00	15.175
k141_11989	29.802	AY278741.1	SARS coronavirus Urbani, complete genome	82,300	26.632	0,000E+00	15.169
k141_7303	16.036	AP019846.1	Leptotrichia hongkongensis JMUB5056 DNA, complete genome	99,183	16.035	0,000E+00	28.884
k141_20796	13.656	LR778174.1	Veillonella parvula strain SKV38 genome assembly, chromosome: 1	99,817	13.653	0,000E+00	25.074
k141_21803	11.776	CP012410.1	Leptotrichia sp, oral taxon 212 strain W10393, complete genome	99,236	11.777	0,000E+00	21.246
k141_19444	8.633		No significant similarity found.				
k141_19747	8.210	CP072359.1	Prevotella melaninogenica strain F0091 chromosome 2, complete sequence	97,381	8.209	0,000E+00	13.967
k141_10767	7.584		No significant similarity found.				
k141_21891	7.432	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	98,145	7.333	0,000E+00	12.787
k141_17952	7.169	CP068257.2	Homo sapiens isolate CHM13 chromosome 21	98,558	5.825	0,000E+00	10.277
k141_17952	7.169	FP236383.1	Human DNA sequence from clone CH507-528H12 on chromosome 21, complete sequence	98,558	5.826	0,000E+00	10.277
k141_17952	7.169	MF164268.1	Homo sapiens clone BAC JH12 genomic sequence	98,558	5.824	0,000E+00	10.275
k141_17952	7.169	MF164265.1	Homo sapiens clone BAC JH6 genomic sequence	98,558	5.824	0,000E+00	10.275
k141_17952	7.169	MF164263.1	Homo sapiens clone BAC JH3 genomic sequence	98,558	5.824	0,000E+00	10.275
k141_798	7.150	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	97,929	7.147	0,000E+00	12.379
k141_19768	7.024	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	93,910	7.028	0,000E+00	10.599
k141_13219	5.924	MZ092489.1	Homo sapiens isolate 36536940 LINE 1, complete sequence	98,992	5.856	0,000E+00	10.484
k141_13219	5.924	MZ092301.1	Homo sapiens isolate 69805717 LINE 1, complete sequence	98,890	5.856	0,000E+00	10.447
k141_13219	5.924	AC008496.6	Homo sapiens chromosome 5 clone CTC-431G16, complete sequence	98,889	5.853	0,000E+00	10.447
k141_13219	5.924	MZ092335.1	Homo sapiens isolate 157547718 LINE 1, complete sequence	98,873	5.856	0,000E+00	10.447
k141_13219	5.924	CP034499.1	Eukaryotic synthetic construct chromosome 20	98,824	5.865	0,000E+00	10.445
k141_2458	5.854	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	95,914	5.825	0,000E+00	9.415
k141_6987	5.802		No significant similarity found.				
k141_5255	5.659	CP040006.1	Schaalia odontolytica strain XH001 chromosome, complete genome	96,364	5.115	0,000E+00	8.379
k141_12253	5.414	CP012410.1	Leptotrichia sp, oral taxon 212 strain W10393, complete genome	96,229	4.932	0,000E+00	8.059
k141_10094	5.338	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	98,782	5.336	0,000E+00	9.492
k141_10156	5.173	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	97,969	5.169	0,000E+00	8.964
k141_10855	4.989	CP072345.1	Prevotella melaninogenica strain F0692 chromosome 1, complete sequence	95,516	2.788	0,000E+00	4.457
k141_17480	4.869	CP072331.1	Prevotella veroralis strain F0319 chromosome 2, complete sequence	98,122	4.792	0,000E+00	8.346
k141_12356	4.772	LR134384.1	Prevotella oris strain NCTC13071 genome assembly, chromosome: 1	87,967	4.712	0,000E+00	5.542
k141_19351	4.750	CP065990.1	Veillonella parvula strain FDAARGOS_1046 chromosome, complete genome	95,833	48	2,130E-08	77
k141_16413	4.712	CP001650.1	Zunongwangia profunda SM-A87, complete genome	100,000	29	9,900E-02	55
k141_5355	4.708	LR778174.1	Veillonella parvula strain SKV38 genome assembly, chromosome: 1	99,745	4.708	0,000E+00	8.628
k141_27644	4.414	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	97,281	4.413	0,000E+00	7.483
k141_23318	4.389	CP072331.1	Prevotella veroralis strain F0319 chromosome 2, complete sequence	98,313	4.387	0,000E+00	7.692
k141_12081	4.300	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	97,836	4.298	0,000E+00	7.422
k141_3691	4.291	LR778174.1	Veillonella parvula strain SKV38 genome assembly, chromosome: 1	99,790	4.291	0,000E+00	7.875

Tabelle 2: *Die 50 längsten zusammengesetzten Contigs mit Megahit (v.1.2.9), Datenbankabfrage vom 20.12.2021*

Contig ID	Contig length	GenBank accession	Subect title	Identity (%)	Length	E-value	Bitscore
k141_3454	4.225		No significant similarity found.				
k141_9806	4.222	CP072364.1	Prevotella jejuni strain F0697 chromosome 2, complete sequence	98,982	4.222	0,000E+00	7.559
k141_7015	4.133	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	98,692	4.130	0,000E+00	7.328
k141_7976	4.088	CP072365.1	Prevotella jejuni strain F0106 chromosome 1, complete sequence	98,017	4.085	0,000E+00	7.095
k141_5351	4.084	OU452294.1	Pammene fasciana genome assembly, chromosome: 22	94,444	36	8,600E-02	55
k141_16288	4.068	MZ824237.1	Reagent-associated CRESS-like virus 1 isolate 7 replicase-like gene, partial sequence	99,691	3.878	0,000E+00	7.095
k141_14229	4.054	CP072331.1	Prevotella veroralis strain F0319 chromosome 2, complete sequence	98,448	3.995	0,000E+00	7.031
k141_25910	4.028	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	98,559	4.025	0,000E+00	7.112
k141_1855	3.995	CP072331.1	Prevotella veroralis strain F0319 chromosome 2, complete sequence	97,321	3.994	0,000E+00	6.780
k141_18965	3.971	XM_045317038.1	PREDICTED: Mercenaria mercenaria uncharacterized LOC123534692 (LOC123534692), mRNA	100,000	29	8,300E-02	55
k141_11609	3.889	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	98,277	3.889	0,000E+00	6.811
k141_7725	3.861		No significant similarity found.				
k141_4971	3.759	CP019721.1	Veillonella parvula strain UTDB1-3, complete genome	98,857	3.761	0,000E+00	6.706
k141_8774	3.756	AP019846.1	Leptotrichia hongkongensis JMUB5056 DNA, complete genome	97,815	3.753	0,000E+00	6.473
k141_21871	3.723	CP072331.1	Prevotella veroralis strain F0319 chromosome 2, complete sequence	100,000	30	2,200E-02	57
k141_4917	3.703	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	97,743	3.678	0,000E+00	6.329
k141_12306	3.688	MW046375.1	Phoenicopteridae parvo-like hybrid virus isolate par083par024 genomic sequence	94,259	3.292	0,000E+00	5.018
k141_3180	3.634		No significant similarity found.				
k141_24646	3.624	CP072331.1	Prevotella veroralis strain F0319 chromosome 2, complete sequence	98,592	3.623	0,000E+00	6.408
k141_10404	3.536	CP019721.1	Veillonella parvula strain UTDB1-3, complete genome	99,491	3.536	0,000E+00	6.429
k141_14387	3.526	CP003667.1	Prevotella sp, oral taxon 299 str, F0039 plasmid, complete sequence	93,381	2.100	0,000E+00	3.083
k141_5501	3.498	CP072330.1	Prevotella veroralis strain F0319 chromosome 1, complete sequence	97,368	3.495	0,000E+00	5.945

Name	GenBank accession number	Length (nt)	RNA-Source
SARS-CoV-2	MN908947.3	29.903	Bronchoalveolar lavage fluid (human)
Bat-SARS-CoV	MG772933.1	29.802	Intestinal tissues (bat)
Bat-SARS-CoV_Short	MG772933_short*	6.420	
HIV	LC312715.1	8.819	Stocked plasma samples (human)
Hepatitis delta	NC_001653.2	1.682	Serum (human)
Zika	NC_035889.1	10.808	Placenta, lungs, heart, skin, spleen, thymus, liver, kidneys, and cerebral cortex (human)
Measles 1	AF266291.1	15.894	Vero cells
Measles 2	KJ410048.1	15.894	Throat swab (human), Vero-hSLAM cells
SARS-CoV	AY278741.1	29.727	Throat washings (human), Vero cells
SARS-tor	AY274119.3	29.751	
Ebola	NC_039345.1	19.043	Oral and rectal swabs, and whole blood (animal)
Marburg	NC_024781.1	19.114	Blood and serum specimens, Vero E6 cells
Rnd-Uniform	rnd_uniform*	29.903	random simulation
Rnd-Wuhan	rnd_wuhan*	29.903	random simulation
Rnd-MK-1	rnd_wh_mk_1*	29.903	random simulation
Rnd-MK-2	rnd_wh_mk_2*	29.903	random simulation

*) Name FASTA-Dateien.

Tabelle 3: *In dieser Studie verwendete Referenzsequenzen.*

ID	Library	Name	Accession	Length	Number of mapped reads	Minimum length in % (M1)	Minimum identity in % (M2)	Number selected reads	Proportion of mapped reads	Number Contigs	Longest contig	Error rate in % related to to the contigs (R1)	related to to reference sequence length (R2)
1	SRR10971381	SARS-CoV-2	MN908947	29.903	479.694	47	0,50	264.281	55,09%	1	29.903	0,00%	0,00%
2	SRR10971381	SARS-CoV-2	MN908947	29.903	479.694	37	0,60	227.010	47,32%	1	29.903	0,00%	0,00%
3	SRR10971381	SARS-CoV-2	MN908947	29.903	479.694	32	0,60	262.885	54,80%	1	29.903	0,00%	0,00%
4	SRR10971381	SARS-CoV-2	MN908947	29.903	479.694	30	0,60	270.025	56,29%	1	29.903	0,00%	0,00%
5	SRR10971381	SARS-CoV-2	MN908947	29.903	479.694	25	0,62	233.537	48,68%	1	29.903	0,00%	0,00%
6	SRR10971381	SARS-CoV-2	MN908947	29.903	479.694	47 (max. 100)	0,50	131.893	27,50%	1	29.855	19,10%	19,23%
7	SRR10971381	SARS-CoV-2	MN908947	29.903	479.694	37 (max. 100)	0,60	98.776	20,59%	1	29.878	29,90%	29,96%
8	SRR10971381	SARS-CoV-2	MN908947	29.903	479.694	32 (max. 100)	0,60	134.651	28,07%	1	29.885	16,00%	16,05%
9	SRR10971381	SARS-CoV-2	MN908947	29.903	479.694	30 (max. 100)	0,60	141.791	29,56%	1	29.893	14,10%	14,13%
10	SRR10971381	SARS-CoV-2	MN908947	29.903	479.694	25 (max. 100)	0,62	105.604	22,01%	1	29.838	28,60%	28,76%
11	SRR10971381	Bat-SARS-CoV	MG772933	29.802	493.888	47	0,50	274.167	55,51%	1	29.802	9,70%	9,70%
12	SRR10971381	Bat-SARS-CoV	MG772933	29.802	493.888	37	0,60	227.731	46,11%	1	29.802	10,20%	10,20%
13	SRR10971381	Bat-SARS-CoV	MG772933	29.802	493.888	32	0,60	263.686	53,39%	1	29.802	9,70%	9,70%
14	SRR10971381	Bat-SARS-CoV	MG772933	29.802	493.888	30	0,60	270.827	54,84%	1	29.802	9,60%	9,60%
15	SRR10971381	Bat-SARS-CoV	MG772933	29.802	493.888	25	0,62	234.553	47,49%	1	29.802	10,20%	10,20%
16	SRR10971381	Bat-SARS-CoV_Short	MG772933_short	6.420	197.266	47	0,50	64.677	32,79%	1	6.410	11,70%	11,84%
17	SRR10971381	Bat-SARS-CoV_Short	MG772933_short	6.420	197.266	37	0,60	68.358	34,65%	1	6.414	12,20%	12,28%
18	SRR10971381	Bat-SARS-CoV_Short	MG772933_short	6.420	197.266	32	0,60	92.334	46,81%	1	6.420	10,70%	10,70%
19	SRR10971381	Bat-SARS-CoV_Short	MG772933_short	6.420	197.266	30	0,60	97.431	49,39%	1	6.420	10,60%	10,60%
20	SRR10971381	Bat-SARS-CoV_Short	MG772933_short	6.420	197.266	25	0,62	81.035	41,08%	1	6.420	12,00%	12,00%
21	SRR10971381	HIV	LC312715.1	8.819	315.060	47	0,50	125.861	39,95%	1	8.802	8,10%	8,28%
22	SRR10971381	HIV	LC312715.1	8.819	315.060	37	0,60	105.035	33,34%	1	8.797	8,60%	8,83%
23	SRR10971381	HIV	LC312715.1	8.819	315.060	32	0,60	140.425	44,57%	1	8.811	2,40%	2,49%
24	SRR10971381	HIV	LC312715.1	8.819	315.060	30	0,60	147.702	46,88%	1	8.814	2,00%	2,06%
25	SRR10971381	HIV	LC312715.1	8.819	315.060	25	0,62	112.543	35,72%	1	8.814	7,60%	7,65%
26	SRR10971381	Hepatitis Delta	NC_001653.2	1.682	163.002	47	0,50	59.234	36,34%	1	1.647	4,00%	6,00%
27	SRR10971381	Hepatitis Delta	NC_001653.2	1.682	163.002	37	0,60	49.517	30,38%	1	1.656	7,10%	8,54%
28	SRR10971381	Hepatitis Delta	NC_001653.2	1.682	163.002	32	0,60	70.689	43,37%	1	1.677	2,30%	2,59%
29	SRR10971381	Hepatitis Delta	NC_001653.2	1.682	163.002	30	0,60	74.744	45,85%	1	1.677	1,70%	1,99%
30	SRR10971381	Hepatitis Delta	NC_001653.2	1.682	163.002	25	0,62	60.074	36,85%	1	1.677	4,80%	5,08%
31	SRR10971381	Zika	NC_035889.1	10.808	310.070	47	0,50	105.438	34,00%	1	10.759	14,40%	14,79%
32	SRR10971381	Zika	NC_035889.1	10.808	310.070	37	0,60	87.258	28,14%	1	10.767	17,30%	17,61%
33	SRR10971381	Zika	NC_035889.1	10.808	310.070	32	0,60	126.216	40,71%	1	10.802	5,70%	5,75%
34	SRR10971381	Zika	NC_035889.1	10.808	310.070	30	0,60	133.190	42,95%	1	10.802	4,70%	4,75%
35	SRR10971381	Zika	NC_035889.1	10.808	310.070	25	0,62	100.930	32,55%	1	10.789	13,60%	13,75%
36	SRR10971381	Measles 1	AF266291.1	15.894	313.628	47	0,50	100.344	31,99%	1	15.818	24,40%	24,76%
37	SRR10971381	Measles 1	AF266291.1	15.894	313.628	37	0,60	87.565	27,92%	1	15.872	29,20%	29,30%
38	SRR10971381	Measles 1	AF266291.1	15.894	313.628	32	0,60	123.325	39,32%	1	15.886	11,60%	11,64%
39	SRR10971381	Measles 1	AF266291.1	15.894	313.628	30	0,60	130.537	41,62%	1	15.886	9,40%	9,45%
40	SRR10971381	Measles 1	AF266291.1	15.894	313.628	25	0,62	99.632	31,77%	1	15.881	24,20%	24,26%
41	SRR10971381	Measles 2	KJ410048.1	15.894	304.700	47	0,50	94.216	30,92%	1	15.837	24,80%	25,07%
42	SRR10971381	Measles 2	KJ410048.1	15.894	304.700	37	0,60	85.392	28,02%	1	15.875	28,70%	28,79%
43	SRR10971381	Measles 2	KJ410048.1	15.894	304.700	32	0,60	120.754	39,63%	1	15.886	11,40%	11,44%
44	SRR10971381	Measles 2	KJ410048.1	15.894	304.700	30	0,60	127.848	41,96%	1	15.885	9,50%	9,55%
45	SRR10971381	Measles 2	KJ410048.1	15.894	304.700	25	0,62	96.089	31,54%	1	15.841	24,50%	24,75%

Tabelle 4: *Ergebnisse der Konsenssequenzanalysen.*

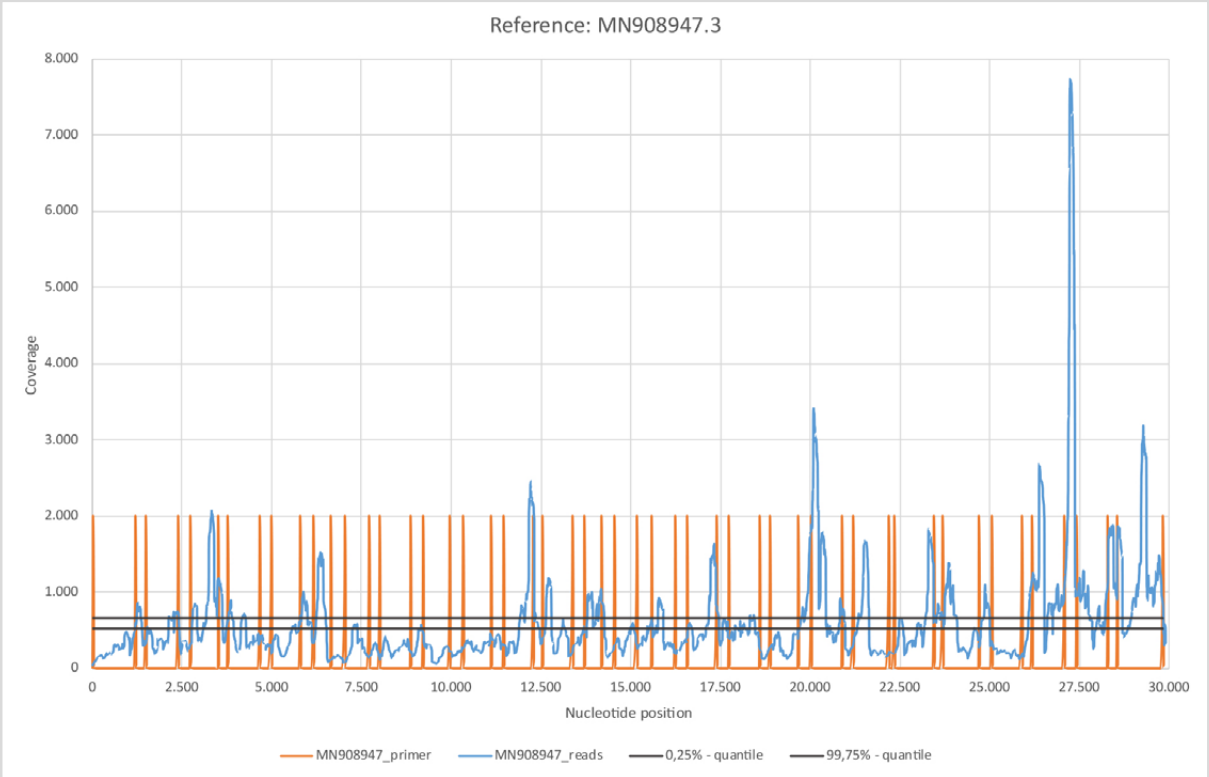
ID	Library	Name	Accession	Length	Number of mapped reads	Minimum length in % (M1)	Minimum identity in % (M2)	Number selected reads	Proportion of mapped reads	Number Contigs	Longest contig	Error rate in % related to to the contigs (R1)	Error rate in % related to to reference sequence length (R2)
46	SRR10971381	SARS-CoV	AY278741.1	29.727	460.238	47	0,50	238.799	51,89%	1	29.727	11,30%	11,30%
47	SRR10971381	SARS-CoV	AY278741.1	29.727	460.238	37	0,60	210.410	45,72%	1	29.727	12,80%	12,80%
48	SRR10971381	SARS-CoV	AY278741.1	29.727	460.238	32	0,60	244.768	53,18%	1	29.727	11,30%	11,30%
49	SRR10971381	SARS-CoV	AY278741.1	29.727	460.238	30	0,60	251.774	54,71%	1	29.727	11,10%	11,10%
50	SRR10971381	SARS-CoV	AY278741.1	29.727	460.238	25	0,62	215.316	46,78%	1	29.727	12,70%	12,70%
51	SRR10971381	SARS-tor	AY274119.3	29.751	462.518	47	0,50	240.919	52,09%	1	29.751	11,30%	11,30%
52	SRR10971381	SARS-tor	AY274119.3	29.751	462.518	37	0,60	212.884	46,03%	1	29.751	12,80%	12,80%
53	SRR10971381	SARS-tor	AY274119.3	29.751	462.518	32	0,60	247.370	53,48%	1	29.751	11,30%	11,30%
54	SRR10971381	SARS-tor	AY274119.3	29.751	462.518	30	0,60	254.412	55,01%	1	29.751	11,10%	11,10%
55	SRR10971381	SARS-tor	AY274119.3	29.751	462.518	25	0,62	218.011	47,14%	1	29.751	12,80%	12,80%
56	SRR10971381	Ebola	NC_039345.1	19.043	307.532	47	0,50	94.263	30,65%	1	19.043	26,80%	26,80%
57	SRR10971381	Ebola	NC_039345.1	19.043	307.532	37	0,60	84.338	27,42%	5	15.130	30,40%	–
58	SRR10971381	Ebola	NC_039345.1	19.043	307.532	32	0,60	118.026	38,38%	1	19.043	16,20%	16,20%
59	SRR10971381	Ebola	NC_039345.1	19.043	307.532	30	0,60	125.008	40,65%	1	19.043	13,80%	13,80%
60	SRR10971381	Ebola	NC_039345.1	19.043	307.532	25	0,62	94.095	30,60%	1	18.808	32,00%	32,84%
61	SRR10971381	Marburg	NC_024781.1	19.114	318.728	47	0,50	111.206	34,89%	1	19.106	19,70%	19,73%
62	SRR10971381	Marburg	NC_024781.1	19.114	318.728	37	0,60	86.021	26,99%	1	19.090	31,60%	31,69%
63	SRR10971381	Marburg	NC_024781.1	19.114	318.728	32	0,60	125.752	39,45%	1	19.108	14,20%	14,23%
64	SRR10971381	Marburg	NC_024781.1	19.114	318.728	30	0,60	133.125	41,77%	1	19.102	12,20%	12,26%
65	SRR10971381	Marburg	NC_024781.1	19.114	318.728	25	0,62	100.037	31,39%	1	19.107	30,20%	30,23%
66	SRR10971381	Rnd-Uniform	rnd_uniform	29.903	333.742	47	0,50	107.283	32,15%	18	3.904	36,30%	–
67	SRR10971381	Rnd-Uniform	rnd_uniform	29.903	333.742	37	0,60	91.049	27,28%	18	4.509	38,60%	–
68	SRR10971381	Rnd-Uniform	rnd_uniform	29.903	333.742	32	0,60	126.231	37,82%	1	29.753	33,90%	34,23%
69	SRR10971381	Rnd-Uniform	rnd_uniform	29.903	333.742	30	0,60	133.205	39,91%	1	29.794	30,50%	30,75%
70	SRR10971381	Rnd-Uniform	rnd_uniform	29.903	333.742	25	0,62	98.636	29,55%	18	4.467	33,20%	–
71	SRR10971381	Rnd-Wuhan	rnd_wuhan	29.903	313.606	47	0,50	104.960	33,47%	19	3.428	32,30%	–
72	SRR10971381	Rnd-Wuhan	rnd_wuhan	29.903	313.606	37	0,60	82.914	26,44%	19	1.691	36,20%	–
73	SRR10971381	Rnd-Wuhan	rnd_wuhan	29.903	313.606	32	0,60	118.051	37,64%	1	28.937	33,60%	35,75%
74	SRR10971381	Rnd-Wuhan	rnd_wuhan	29.903	313.606	30	0,60	125.287	39,95%	1	29.684	30,30%	30,81%
75	SRR10971381	Rnd-Wuhan	rnd_wuhan	29.903	313.606	25	0,62	93.120	29,69%	17	1.470	34,40%	–
76	SRR10971381	Rnd-MK-1	rnd_wh_mk_1	29.903	327.202	47	0,50	112.718	34,45%	1	29.782	32,30%	32,57%
77	SRR10971381	Rnd-MK-1	rnd_wh_mk_1	29.903	327.202	37	0,60	87.402	26,71%	13	3.899	37,70%	–
78	SRR10971381	Rnd-MK-1	rnd_wh_mk_1	29.903	327.202	32	0,60	126.963	38,80%	1	29.850	25,70%	25,83%
79	SRR10971381	Rnd-MK-1	rnd_wh_mk_1	29.903	327.202	30	0,60	134.450	41,09%	1	29.850	22,60%	22,74%
80	SRR10971381	Rnd-MK-1	rnd_wh_mk_1	29.903	327.202	25	0,62	99.369	30,37%	3	16.838	40,10%	–
81	SRR10971381	Rnd-MK-2	rnd_wh_mk_2	29.903	328.524	47	0,50	116.586	35,49%	1	29.793	34,10%	34,34%
82	SRR10971381	Rnd-MK-2	rnd_wh_mk_2	29.903	328.524	37	0,60	91.824	27,95%	19	4.856	35,30%	–
83	SRR10971381	Rnd-MK-2	rnd_wh_mk_2	29.903	328.524	32	0,60	130.868	39,84%	1	29.881	27,20%	27,25%
84	SRR10971381	Rnd-MK-2	rnd_wh_mk_2	29.903	328.524	30	0,60	138.261	42,09%	1	29.881	24,20%	24,26%
85	SRR10971381	Rnd-MK-2	rnd_wh_mk_2	29.903	328.524	25	0,62	103.219	31,42%	16	8.146	35,20%	–

Software	Commands	Version	Analysis	Citation
SRA Toolkit	`fastq-dump split-files --origfmt --gzip SRR10971381`	2.8.0	Download SRA files	(SRA Toolkit Development Team, no date)
Fastp	`fastp -i SRR10971381_1.fastq.gz -I SRR10971381_2.fastq.gz -o SRR10971381_1.fastq -O SRR10971381_2.fastq`	0.23.1	FASTQ Preprocessing	(Chen et. al., 2018)
Megahit	`megahit -1 SRR10971381_1.fastq -2 SRR10971381_2.fastq -o megahit_result`	1.2.9	De novo Assembly	(Li et al. , 2015)
BBMap	`mapPacBio.sh in=SRR10971381_1.fastq in2=SRR10971381_2.fastq outm=mapped.sam vslow k=8 maxindel=0 minratio=0.1`	38.93	Alignment of short reads	(Bushnell, 2014)
BBMap	`reformat.sh in=../mapped.sam out=sample_selection.sam minlength=$M1 (maxlength=100) idfilter=$M2 ow=t`	38.93	Selection of short reads	(Bushnell, 2014)
BWA	`bwa mem refernce.fasta $left.fastq $right.fastq > out.sam`	0.7.17-r1188	Alignment of short reads	(Li and Durbin, 2013)
Bowtie2	`bowtie2 -x cov -1 SRR10971381_1.fastq -2 SRR10971381_2.fastq --no-unal -p 12 -S sample_final.sam`	2.4.4	Alignment of short reads	(Langmead et. al., 2018)

Samtools	`samtools view -b sample_selection.sam > sample.bam`
	
	`samtools sort sample.bam -o ../sample_sort_reads.bam`
		1.14	Analysis of sam/bam file	(Li et al. , 2009)

	`samtools index ../sample_sort_reads.bam`			
Samtools, bcftools	`samtools mpileup -uf ../mapping/$reference.fasta ../sample_sort_reads.bam | bcftools call -c | vcfutils.pl vcf2fq > SAMPLE_cns.fastq`	1.14	Consensus sequence	(Li et al. , 2009)
Seqtk	`seqtk seq -aQ64 -q20 -n N sample_cns.fastq > sample_cns.fasta`	1.3-r106	Convert to FASTA and quality control	(Li, 2016)

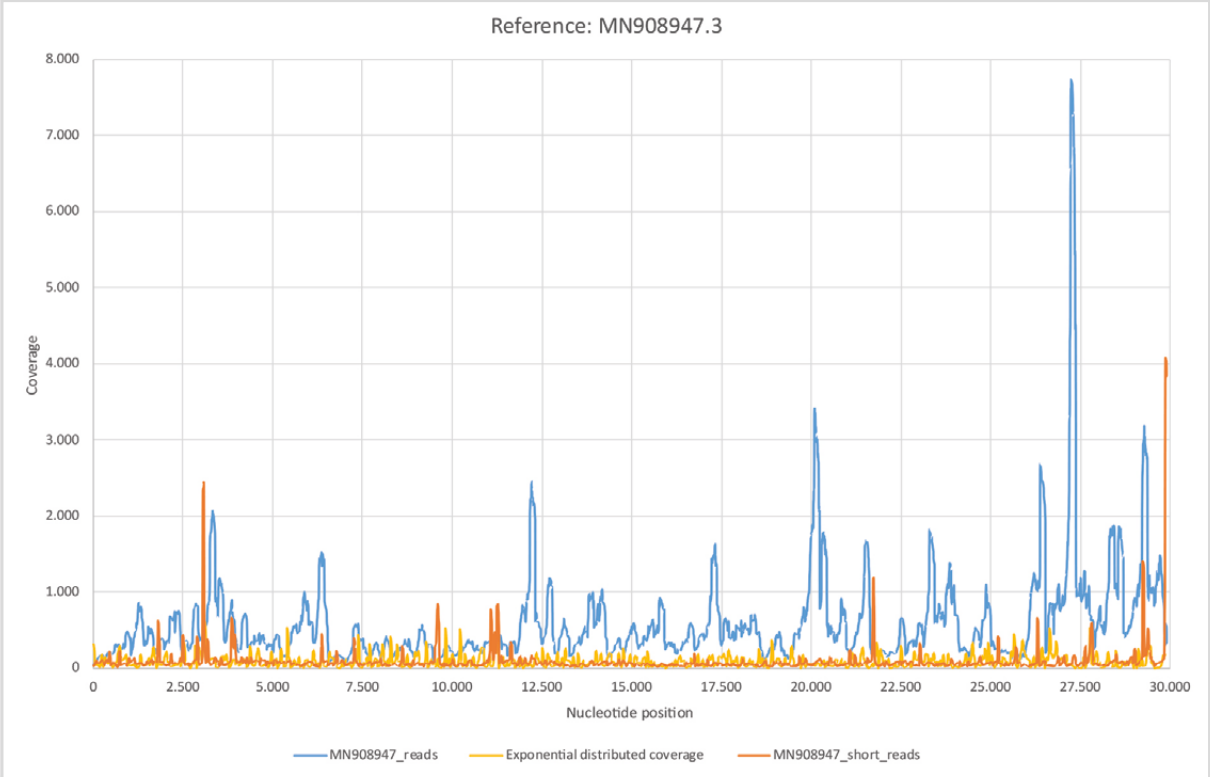
Tabelle 5: *Liste der in dieser Studie verwendeten Software und Befehle.*



Reference - MN908947.3	
Genome length	29.903
Number of reads	121.779
Ø Read length	145,56
P(Covering a nucleotide)	0,00486776
EN (Expected coverage)	592,7907
VARN (Binomial distribution)	589,9052
Covered nucleotides	29.903
Coverage in %	100,00%

Primer	
Genome length	29.903
Number of reads	52
Ø Read length	23,75
P(Covering a nucleotide)	0,00079423
EN (Expected coverage)	0,0413
VARN (Binomial distribution)	0,0413
Covered nucleotides	1.235
Coverage in %	4,13%
Error rate in %	0,00%

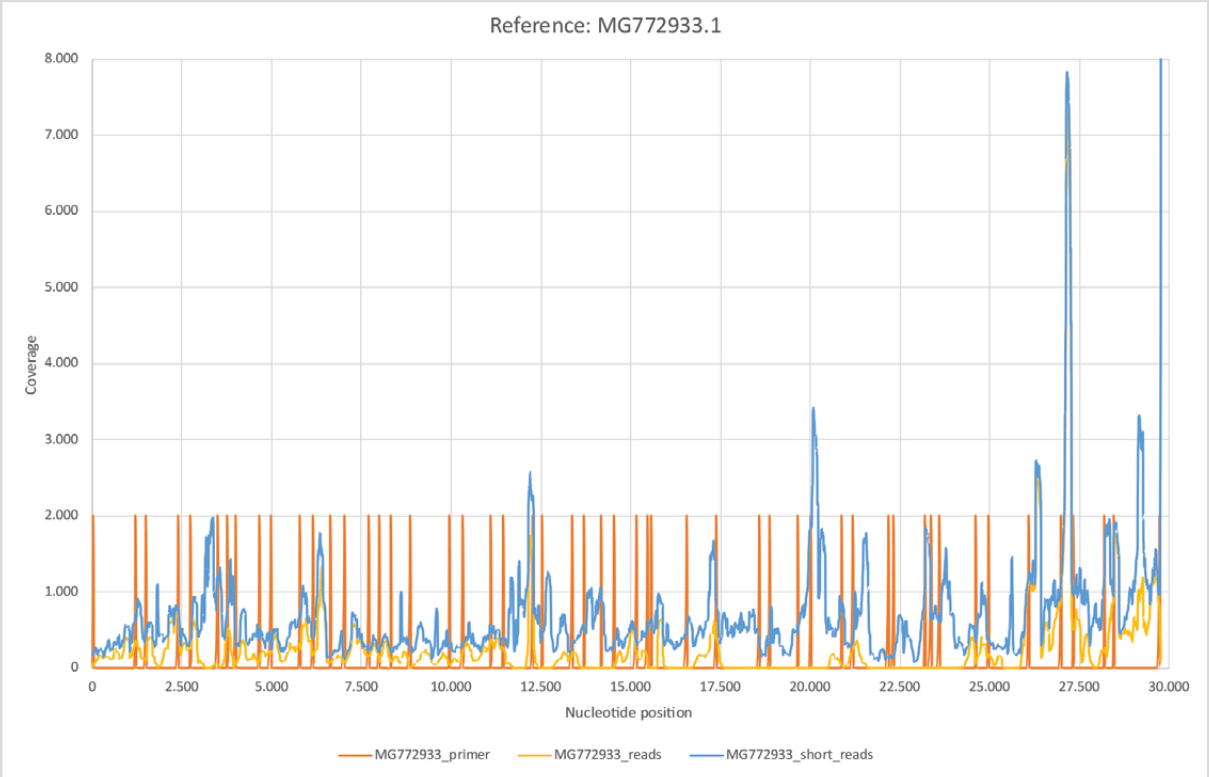
Abbildung 1: *Referenz MN908947.3.*
a) MN908947.3_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) MN908947.3_primer gemappt mit BBMap. c) Quantile wurden ermittelt aus EN und VARN unter der Verteilungshypothese einer Binomialverteilung. d) Die 26 Primerpaare ([1], Supplementary Tabelle 8. PCR primers used in this study.) liegen gleichmäßig verteilt über das gesamte Referenzgenom hinweg. Die Primerpositionen korrelieren mit Bereichen hoher Nukleotidabdeckung.



Reference - MN908947.3	
Genome length	29.903
Number of reads	121.779
Ø Read length	145,56
P(Covering a nucleotide)	0,00486776
EN (Expected coverage)	592,7907
VARN (Binomial distribution)	589,9052
Covered nucleotides	29.903
Coverage in %	100,00%

Reference - MN908947.3 - Short reads	
Genome length	29.903
Number of reads	59.949
Ø Read length	46,24
P(Covering a nucleotide)	0,00154643
Lambda	0,01078668
EN (Expected coverage)	92,7070
VARN (Exponential distribution)	8.595
VARN (Trimmed 99,5%)	19.129
Covered nucleotides	29.903
Coverage in %	100,00%

Abbildung 2: *Referenz MN908947.3.*
a) MN908947.3_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) MN908947.3_short_reads gemappt mit BBMap (M1; M2) = (37 (max. 100); 0,60). c) Exponential distributed coverage wurde durch stochastische Simulation mit Hilfe der Inversionsmethode erzeugt. Die Abdeckungsverteilung MN908947.3_short_reads zeigt ein eher zufälliges Muster, weist aber eine höhere getrimmte Varianz auf. Dies ist im Wesentlichen auf die wenigen Ausschläge der Abdeckungsverteilung zu erklären.

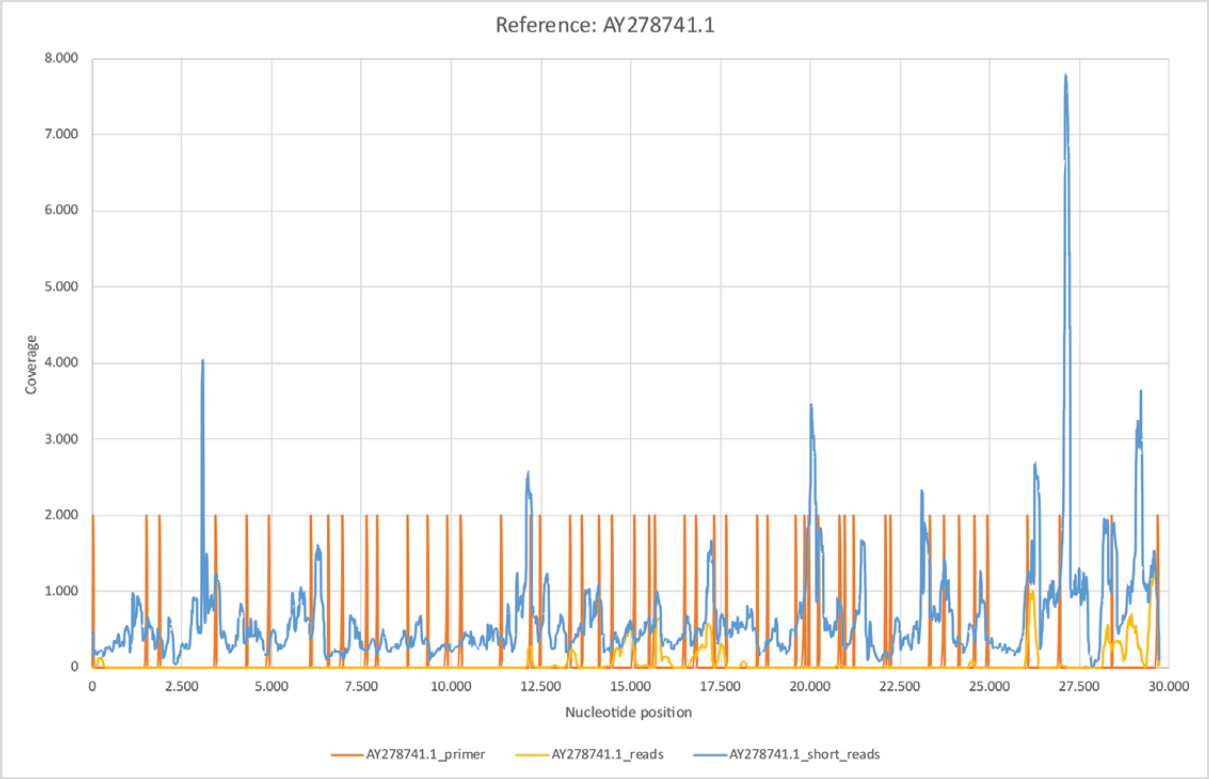


Reference - MG772933.1	
Genome length	29.802
Number of reads	50.722
Ø Read length	146,80
P(Covering a nucleotide)	0,00492572
EN (Expected coverage)	249,8422
VARN (Binomial distribution)	248,6115
Covered nucleotides	22.684
Coverage in %	76,12%

Primer	
Genome length	29.802
Number of reads	52
Ø Read length	23,60
P(Covering a nucleotide)	0,00079176
EN (Expected coverage)	0,0412
VARN (Binomial distribution)	0,0411
Covered nucleotides	1.227
Coverage in %	4,12%
Error rate in %	11,50%

Reference - MG772933.1 - short reads	
Genome length	29.802
Number of reads	183.727
Ø Read length	110,56
P(Covering a nucleotide)	0,00370972
EN (Expected coverage)	681,5748
VARN (Binomial distribution)	679,0464
Covered nucleotides	29.802
Coverage in %	100,00%

Abbildung 3: *Referenz MG772933.1.*
a) MG772933.1_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) MG772933.1_short_reads gemappt mit BBMap (M1; M2) = (37; 0,60). c) MG772933.1_primer mit BBMap. d) Die Abdeckungsverteilung unter a) deckt 76,12% der Referenzsequenz MG772933.1 ab. Mit den Sequenzen unter b) wird eine vollständige Abdeckung erreicht. Die Fehlerraten der berechneten Konsensussequenzen (Tabelle 4, 11-20) zeigen Fehlerraten von etwa 10% in Übereinstimmung mit [1].

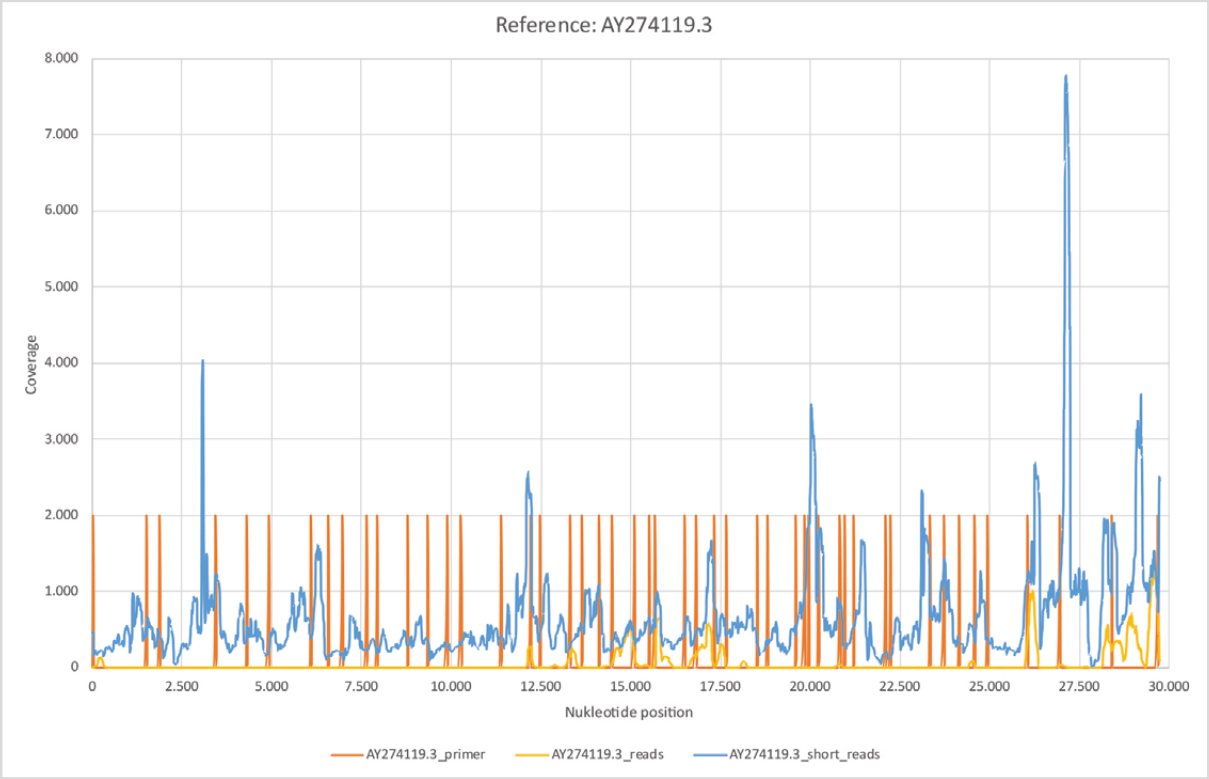


Reference - AY278741.1	
Genome length	29.727
Number of reads	11.332
Ø Read length	145,77
P(Covering a nucleotide)	0,00490377
EN (Expected coverage)	55,5695
VARN (Binomial distribution)	55,2970
Covered nucleotides	8.386
Coverage in %	28,21%

Primer	
Genome length	29.727
Number of reads	52
Ø Read length	23,90
P(Covering a nucleotide)	0,00080411
EN (Expected coverage)	0,0418
VARN (Binomial distribution)	0,0418
Covered nucleotides	1.214
Coverage in %	4,08%
Error rate in %	21,50%

Reference - AY278741.1 - short reads	
Genome length	29.727
Number of reads	168.076
Ø Read length	113,91
P(Covering a nucleotide)	0,00383173
EN (Expected coverage)	644,0224
VARN (Binomial distribution)	641,5547
Covered nucleotides	29.727
Coverage in %	100,00%

Abbildung 4: *Referenz AY278741.1.*
a) AY278741.1_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) AY278741.1_short_reads gemappt mit BBMap (M1; M2) = (37; 0,60). c) AY278741.1_primer mit BBMap. d) Die Abdeckungsverteilung unter a) deckt 28,21% der Referenzsequenz AY278741.1 ab. Mit den Sequenzen unter b) wird eine vollständige Abdeckung erreicht. Die Fehlerraten der berechneten Konsensussequenzen (Tabelle 4, 46–50) zeigen Fehlerraten von etwa 12,8%.

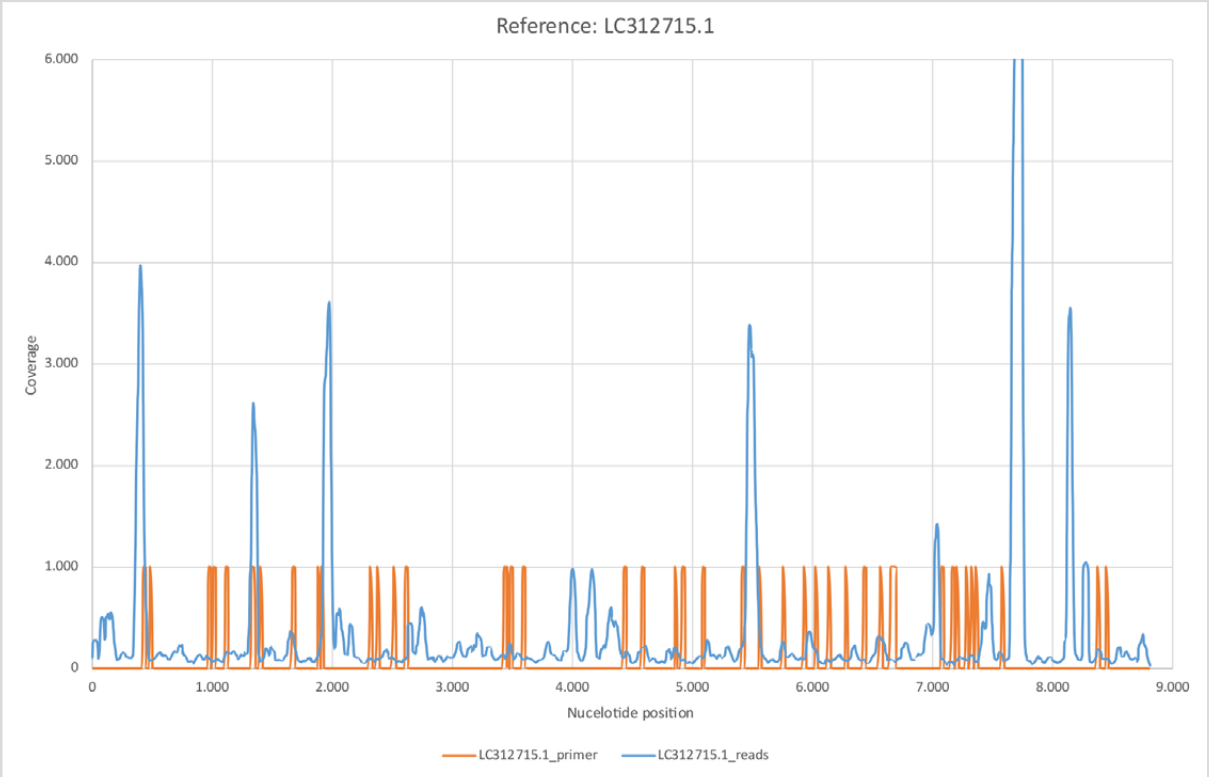


Reference - AY274119.3	
Genome length	29.751
Number of reads	11.419
Ø Read length	144,84
P(Covering a nucleotide)	0,00486833
EN (Expected coverage)	55,5915
VARN (Binomial distribution)	55,3208
Covered nucleotides	8.395
Coverage in %	28,22%

Primer	
Genome length	29.751
Number of reads	52
Ø Read length	23,90
P(Covering a nucleotide)	0,00080346
EN (Expected coverage)	0,0418
VARN (Binomial distribution)	0,0417
Covered nucleotides	1.214
Coverage in %	4,08%
Error rate in %	21,50%

Reference - AY274119.3 - short reads	
Genome length	29.751
Number of reads	170.197
Ø Read length	112,92
P(Covering a nucleotide)	0,00379535
EN (Expected coverage)	645,9568
VARN (Binomial distribution)	643,5051
Covered nucleotides	29.751
Coverage in %	100,00%

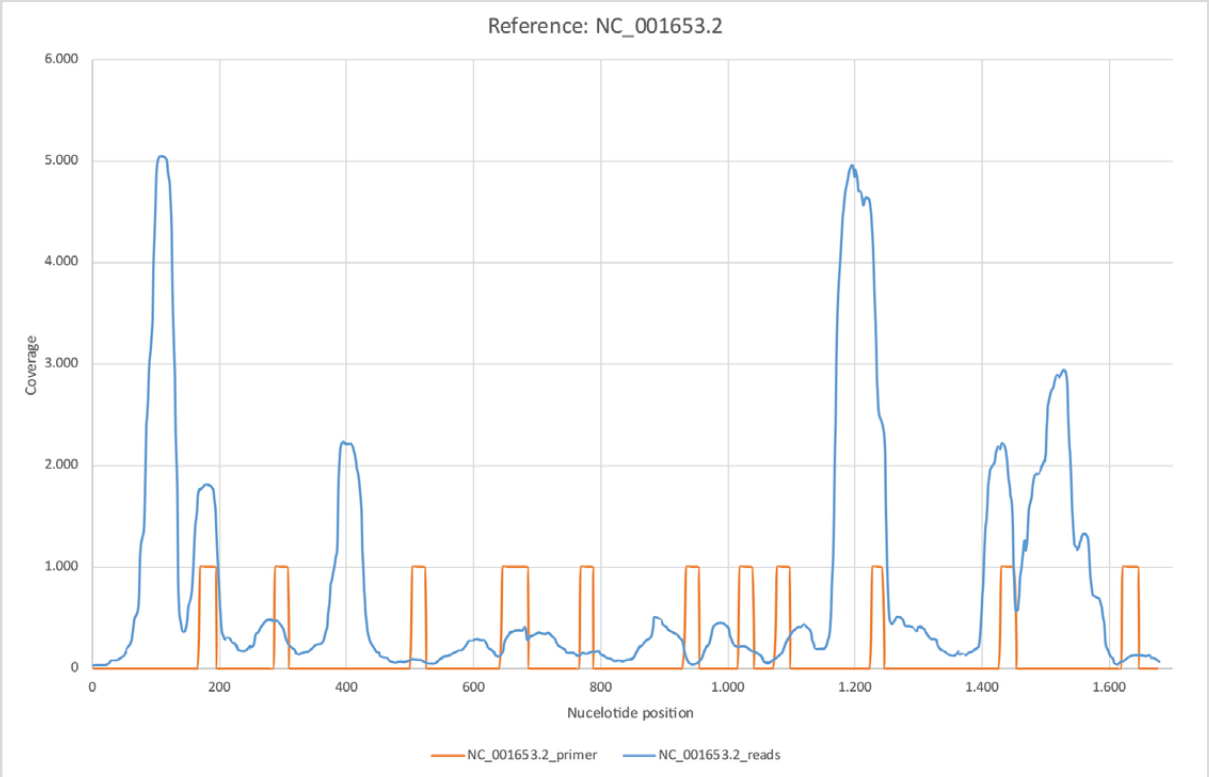
Abbildung 5: *Referenz AY274119.3.*
a) AY274119.3_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) AY274119.3_short_reads gemappt mit BBMap (M1; M2) = (37; 0,60). c) AY274119.3_primer mit BBMap. d) Die Abdeckungsverteilung unter a) deckt 28,22% der Referenzsequenzen AY274119.3 ab. Mit den Sequenzen unter b) wird eine vollständige Abdeckung erreicht. Die Fehlerraten der berechneten Konsensussequenzen (Tabelle 4, 51–55) zeigen Fehlerraten von etwa 12,8%.



Reference - LC312715.1	
Genome length	8.819
Number of reads	65.196
Ø Read length	51,84
P(Covering a nucleotide)	0,00587873
EN (Expected coverage)	383,2696
VARN (Binomial distribution)	381,0165
Covered nucleotides	8.819
Coverage in %	100,00%

Primer	
Genome length	8.819
Number of reads	46
Ø Read length	23,54
P(Covering a nucleotide)	0,00266963
EN (Expected coverage)	0,1228
VARN (Binomial distribution)	0,1225
Covered nucleotides	1.031
Coverage in %	11,69%
Error rate in %	38,00%

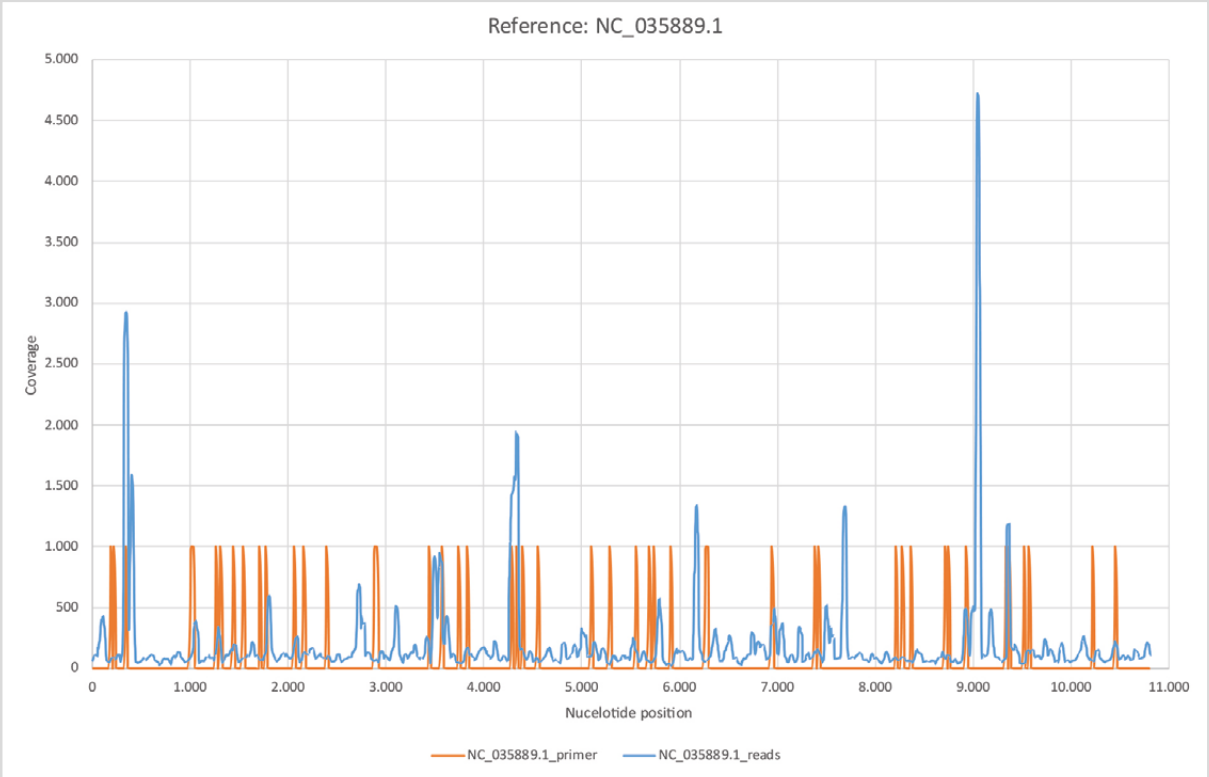
Abbildung 6: *Referenz LC312715.1.*
a) LC312715.1_short_reads gemappt BBMap (M1; M2) = (37; 0,60). b) LC312715.1_primer gemappt mit BBMap.



Reference - NC_001653.2	
Genome length	1.682
Number of reads	27.721
Ø Read length	47,92
P(Covering a nucleotide)	0,02849132
EN (Expected coverage)	789,8080
VARN (Binomial distribution)	767,3053
Covered nucleotides	1.682
Coverage in %	100,00%

Primer	
Genome length	1.682
Number of reads	14
Ø Read length	22,93
P(Covering a nucleotide)	0,01363173
EN (Expected coverage)	0,1908
VARN (Binomial distribution)	0,1882
Covered nucleotides	276
Coverage in %	16,41%
Error rate in %	40,50%

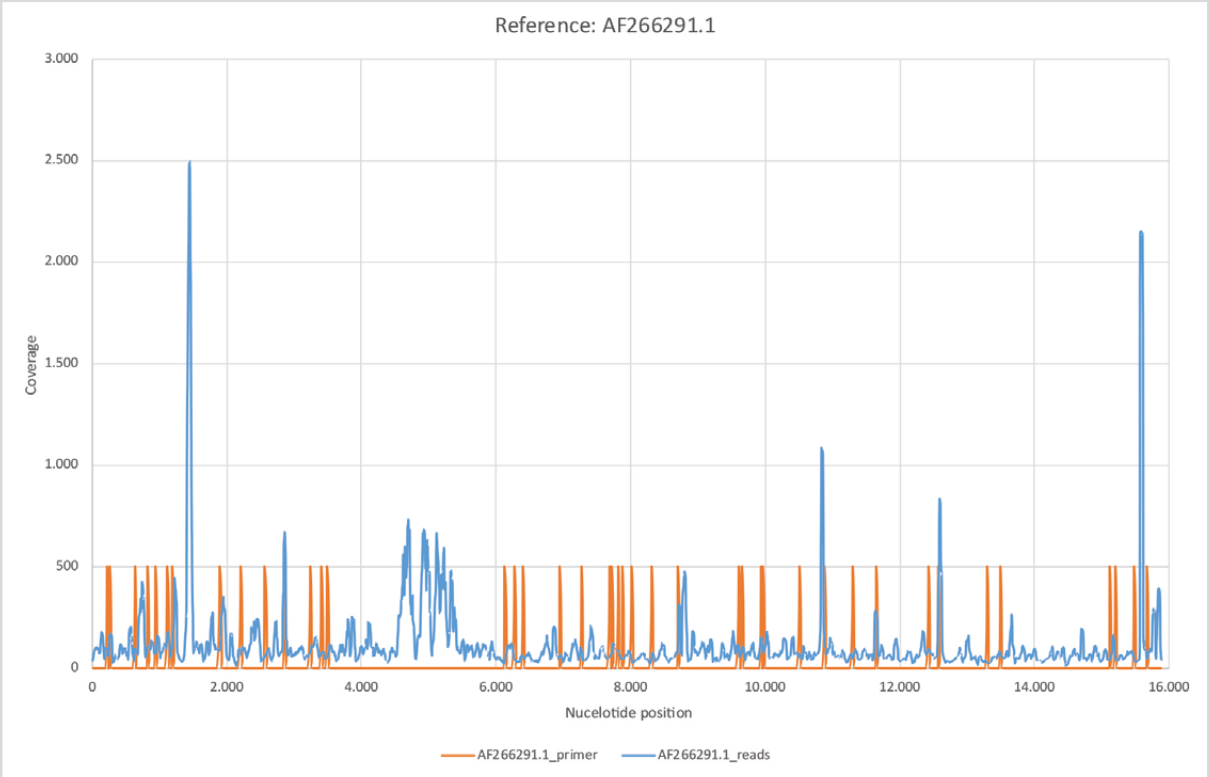
Abbildung 7: *Referenz NC_001653.2.*
a) NC_001653.2_short_reads gemappt BBMap (M1; M2) = (37; 0,60). b) NC_001653.2_primer gemappt mit BBMap.



Reference - NC_035889.1	
Genome length	10.808
Number of reads	48.598
Ø Read length	43,49
P(Covering a nucleotide)	0,00402374
EN (Expected coverage)	195,5456
VARN (Binomial distribution)	194,7588
Covered nucleotides	10.808
Coverage in %	100,00%

Primer	
Genome length	10.808
Number of reads	49
Ø Read length	23,51
P(Covering a nucleotide)	0,00217526
EN (Expected coverage)	0,1066
VARN (Binomial distribution)	0,1064
Covered nucleotides	1.072
Coverage in %	9,92%
Error rate in %	36,90%

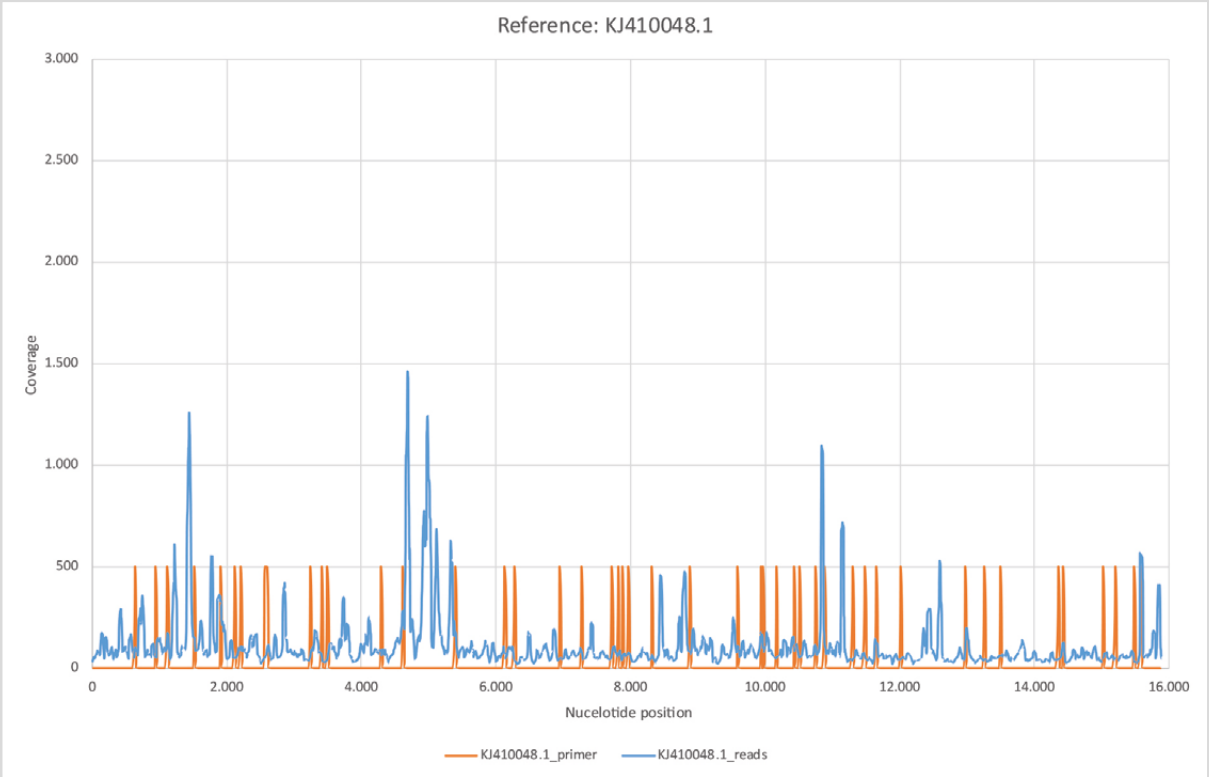
Abbildung 8: *Referenz NC_035889.1.*
a) NC_035889.1_short_reads gemappt BBMap (M1; M2) = (37; 0,60). b) NC_035889.1_primer gemappt mit BBMap.



Reference - AF266291.1	
Genome length	15.894
Number of reads	44.999
Ø Read length	42,58
P(Covering a nucleotide)	0,00267917
EN (Expected coverage)	120,5599
VARN (Binomial distribution)	120,2369
Covered nucleotides	15.894
Coverage in %	100,00%

Primer	
Genome length	15.894
Number of reads	48
Ø Read length	23,56
P(Covering a nucleotide)	0,00148248
EN (Expected coverage)	0,0712
VARN (Binomial distribution)	0,0711
Covered nucleotides	1.053
Coverage in %	6,63%
Error rate in %	36,90%

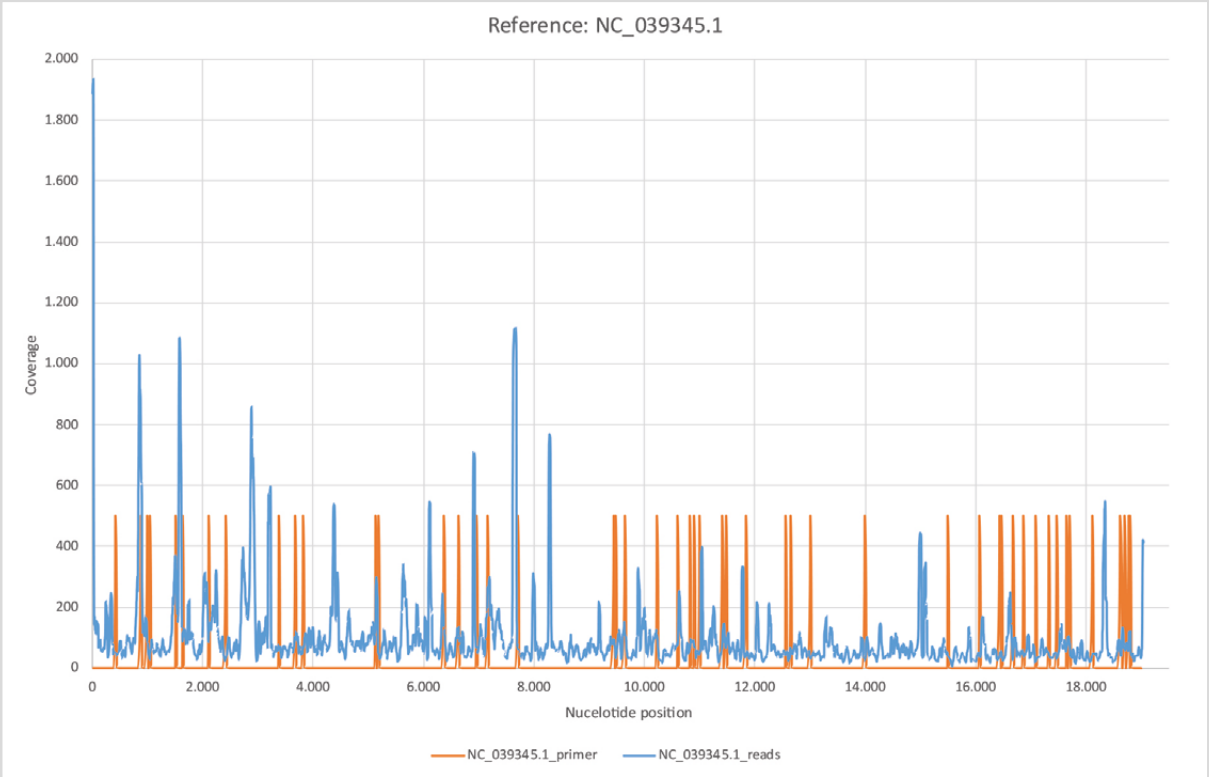
Abbildung 9: *Referenz AF266291.1.*
a) AF266291.1_short_reads gemappt BBMap (M1; M2) = (37; 0,60). b) AF266291.1_primer gemappt mit BBMap.



Reference - KJ410048.1	
Genome length	15.894
Number of reads	42.849
Ø Read length	42,38
P(Covering a nucleotide)	0,00266641
EN (Expected coverage)	114,2528
VARN (Binomial distribution)	113,9482
Covered nucleotides	15.894
Coverage in %	100,00%

Primer	
Genome length	15.894
Number of reads	49
Ø Read length	23,33
P(Covering a nucleotide)	0,00146763
EN (Expected coverage)	0,0719
VARN (Binomial distribution)	0,0718
Covered nucleotides	1.115
Coverage in %	7,02%
Error rate in %	35,10%

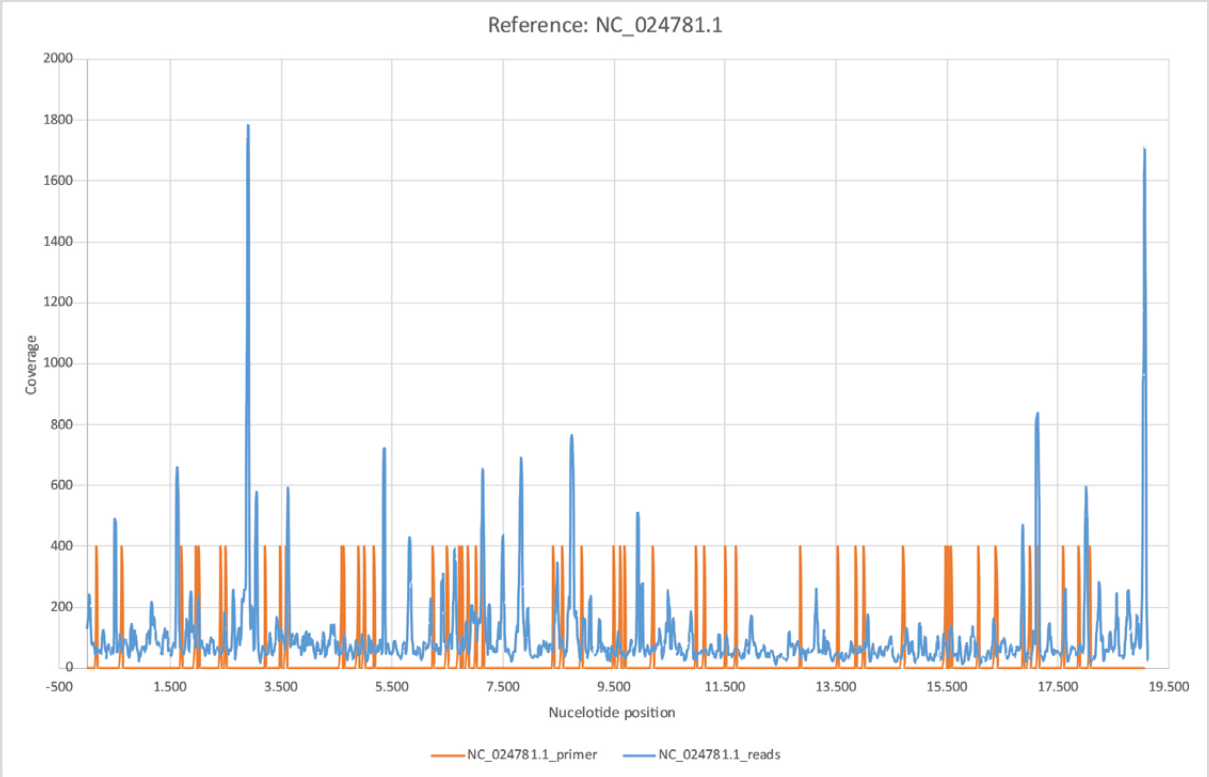
Abbildung 10: *Referenz KJ410048.1.*
a) KJ410048.1_short_reads gemappt BBMap (M1; M2) = (37; 0,60). b) KJ410048.1_primer gemappt mit BBMap.



Reference - NC_039345.1	
Genome length	19.043
Number of reads	46.367
Ø Read length	41,86
P(Covering a nucleotide)	0,00219794
EN (Expected coverage)	101,9121
VARN (Binomial distribution)	101,6881
Covered nucleotides	19.043
Coverage in %	100,00%

Primer	
Genome length	19.043
Number of reads	50
Ø Read length	23,48
P(Covering a nucleotide)	0,00123300
EN (Expected coverage)	0,0616
VARN (Binomial distribution)	0,0616
Covered nucleotides	1.163
Coverage in %	6,11%
Error rate in %	36,20%

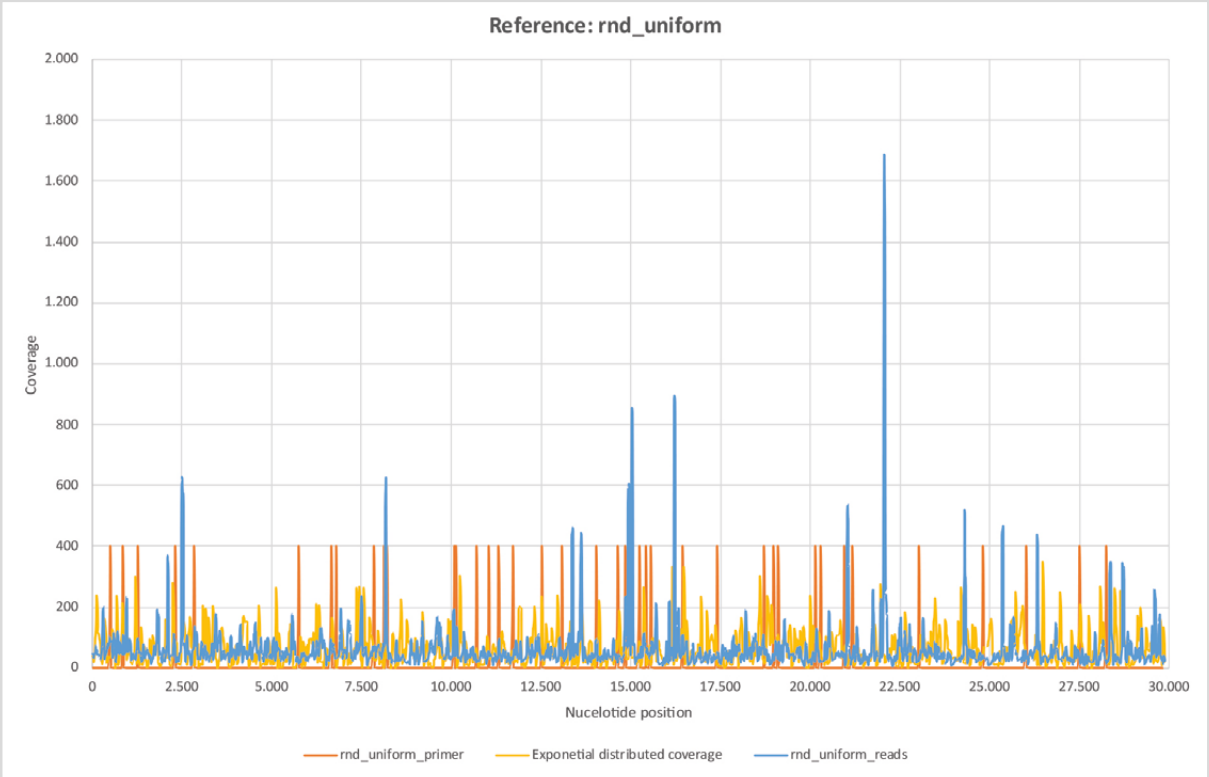
Abbildung 11: *Referenz NC_039345.1.*
a) NC_039345.1_short_reads gemappt BBMap (M1; M2) = (37; 0,60). b) NC_039345.1_primer gemappt mit BBMap.



Reference - NC_024781.1	
Genome length	19.114
Number of reads	45.679
Ø Read length	43,38
P(Covering a nucleotide)	0,00226939
EN (Expected coverage)	103,6633
VARN (Binomial distribution)	103,4281
Covered nucleotides	19.114
Coverage in %	100,00%

Primer	
Genome length	19.114
Number of reads	52
Ø Read length	24,17
P(Covering a nucleotide)	0,00126468
EN (Expected coverage)	0,0658
VARN (Binomial distribution)	0,0657
Covered nucleotides	1.208
Coverage in %	6,32%
Error rate in %	35,80%

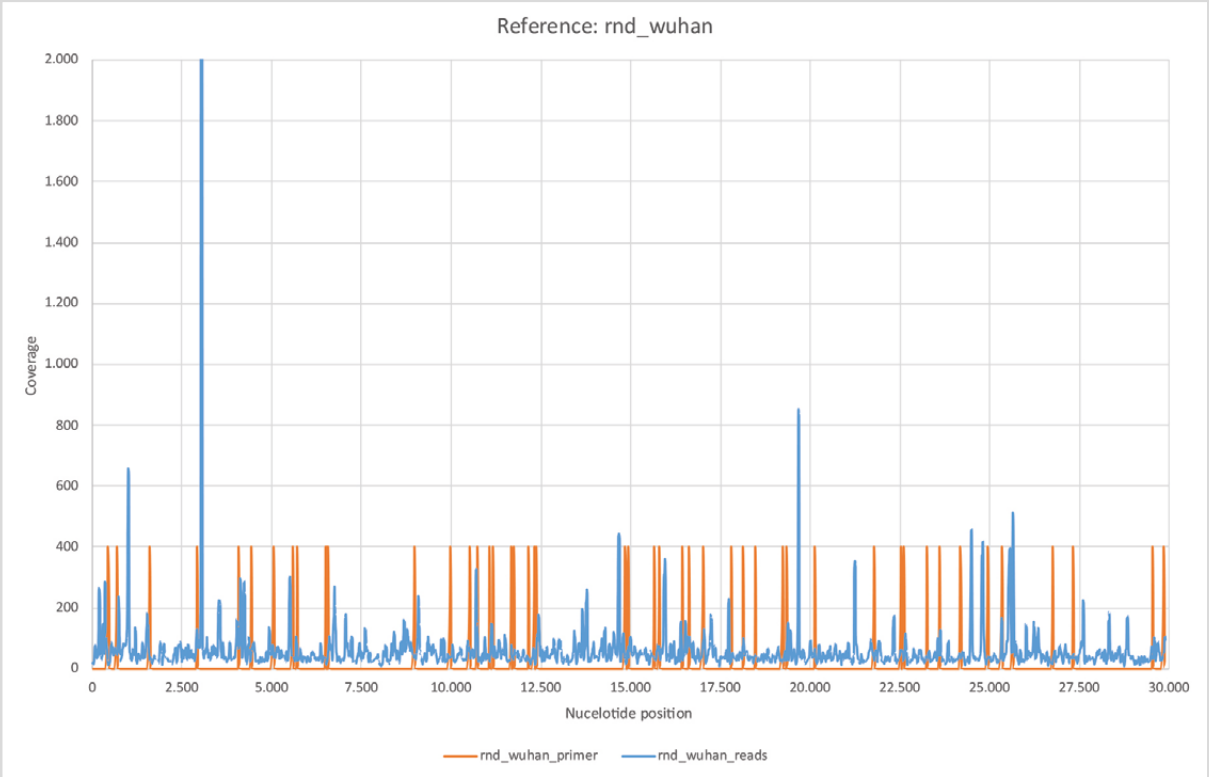
Abbildung 12: *Referenz NC_024781.1.*
a) NC_024781.1_short_reads gemappt BBMap (M1; M2) = (37; 0,60). b) NC_024781.1_primer gemappt mit BBMap.



Reference - rnd_uniform	
Genome length	29.903
Number of reads	46.288
Ø Read length	41,96
P(Covering a nucleotide)	0,00140307
Lambda	0,01539754
EN (Expected coverage)	64,9454
VARN (Exponential distribution)	4.218
VARN (Trimmed 99,5%)	4.125
Covered nucleotides	29.903
Coverage in %	100,00%

Primer	
Genome length	29.903
Number of reads	52
Ø Read length	23,81
P(Covering a nucleotide)	0,00079616
EN (Expected coverage)	0,0414
VARN (Binomial distribution)	0,0414
Covered nucleotides	923
Coverage in %	3,09%
Error rate in %	36,70%

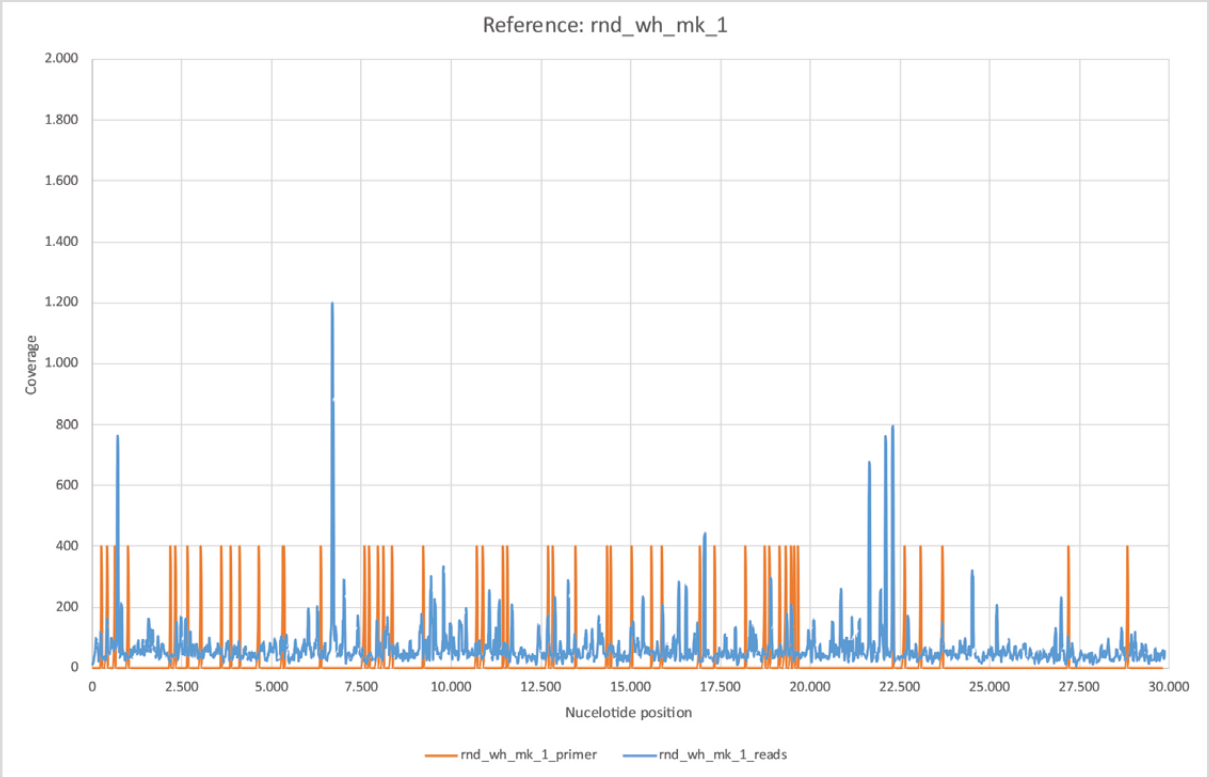
Abbildung 13: *Referenz rnd_uniform.*
a) rnd_uniform_reads gemappt mit BBMap (M1; M2) = (37; 0,60). b) rnd_uniform_primer gemappt mit BBMap.
c) Exponential distributed coverage wurde durch stochastische Simulation mit Hilfe der Inversionsmethode erzeugt.
d) Die 26 Primerpaare ([1, Supplementary Tabelle 8. PCR primers used in this study.]) verteilen sich ungleichmäßig über das gesamte Referenzgenom hinweg. Die Primerpositionen korrelieren teilweise mit Bereichen hoher Nukleotidabdeckung, wobei diese jeweils nur wenige Nukleotide umfassen. e) Die Verteilung von rnd_uniform_reads erscheint weitestgehend zufällig. Die Varianz der betrachteten Exponentialverteilung stimmt gut mit der getrimmten empirischen Varianz überein.



Reference - rnd_wuhan	
Genome length	29.903
Number of reads	44.475
Ø Read length	42,74
P(Covering a nucleotide)	0,00142922
EN (Expected coverage)	63,5644
VARN (Binomial distribution)	63,4736
Covered nucleotides	29.903
Coverage in %	100,00%

Primer	
Genome length	29.903
Number of reads	52
Ø Read length	23,56
P(Covering a nucleotide)	0,00078780
EN (Expected coverage)	0,0410
VARN (Binomial distribution)	0,0409
Covered nucleotides	1.216
Coverage in %	4,07%
Error rate in %	35,00%

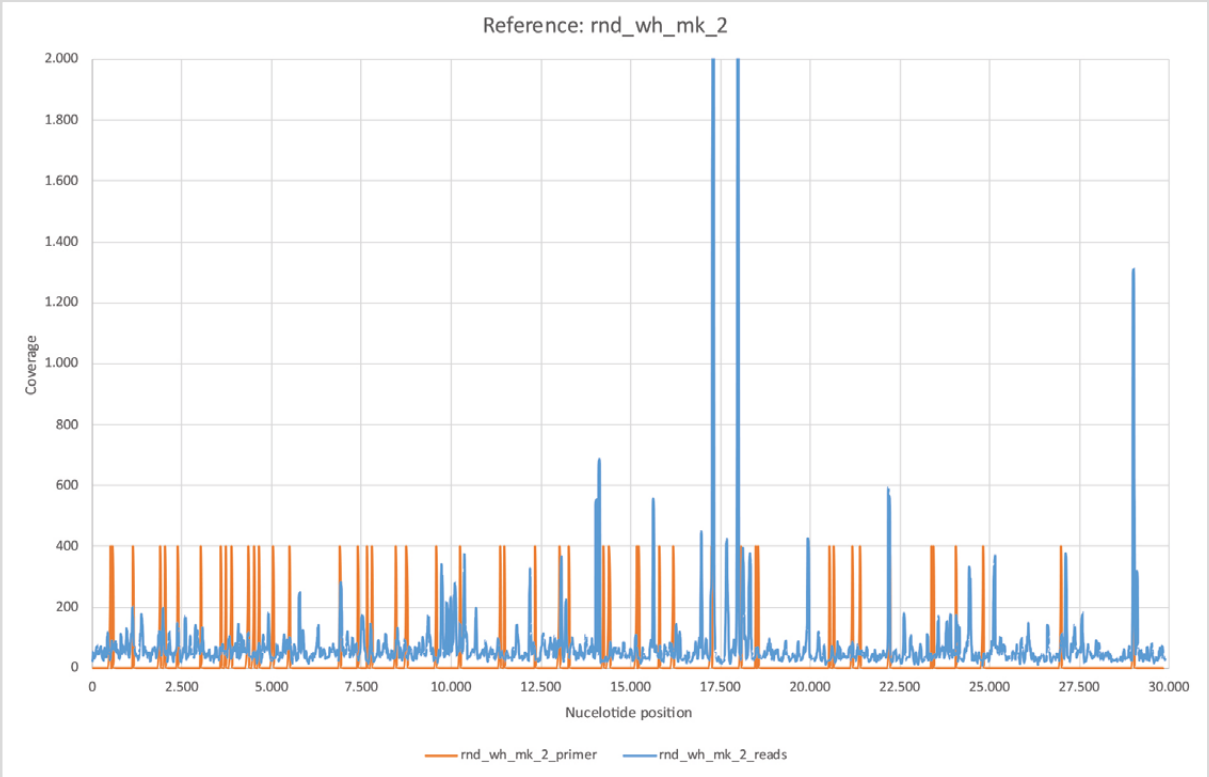
Abbildung 14: *Referenz rnd_wuhan.*
a) rnd_ wuhan _reads gemappt mit BBMap (M1; M2) = (37; 0,60). b) rnd_ wuhan _primer gemappt mit BBMap.
c) Die Abdeckungsverteilung zeigt weitestgehend eine zufällige Verteilung, vergleichbar mit Abbildung 11.
d) Die durchschnittliche Leselänge liegt es oberhalb von rnd_uniform (Abbildung 11). Dies liegt in der verwendeten empirischen Verteilung der Nukleotide (A, T, C und G) gemäß der Referenz für SARS-CoV-2 (MN908947.3).



Reference - rnd_wh_mk_1	
Genome length	29.903
Number of reads	46.779
Ø Read length	42,56
P(Covering a nucleotide)	0,00142314
EN (Expected coverage)	66,5733
VARN (Binomial distribution)	66,4785
Covered nucleotides	29.903
Coverage in %	100,00%

Primer	
Genome length	29.903
Number of reads	50
Ø Read length	23,56
P(Covering a nucleotide)	0,00078788
EN (Expected coverage)	0,0394
VARN (Binomial distribution)	0,0394
Covered nucleotides	1.161
Coverage in %	3,88%
Error rate in %	34,70%

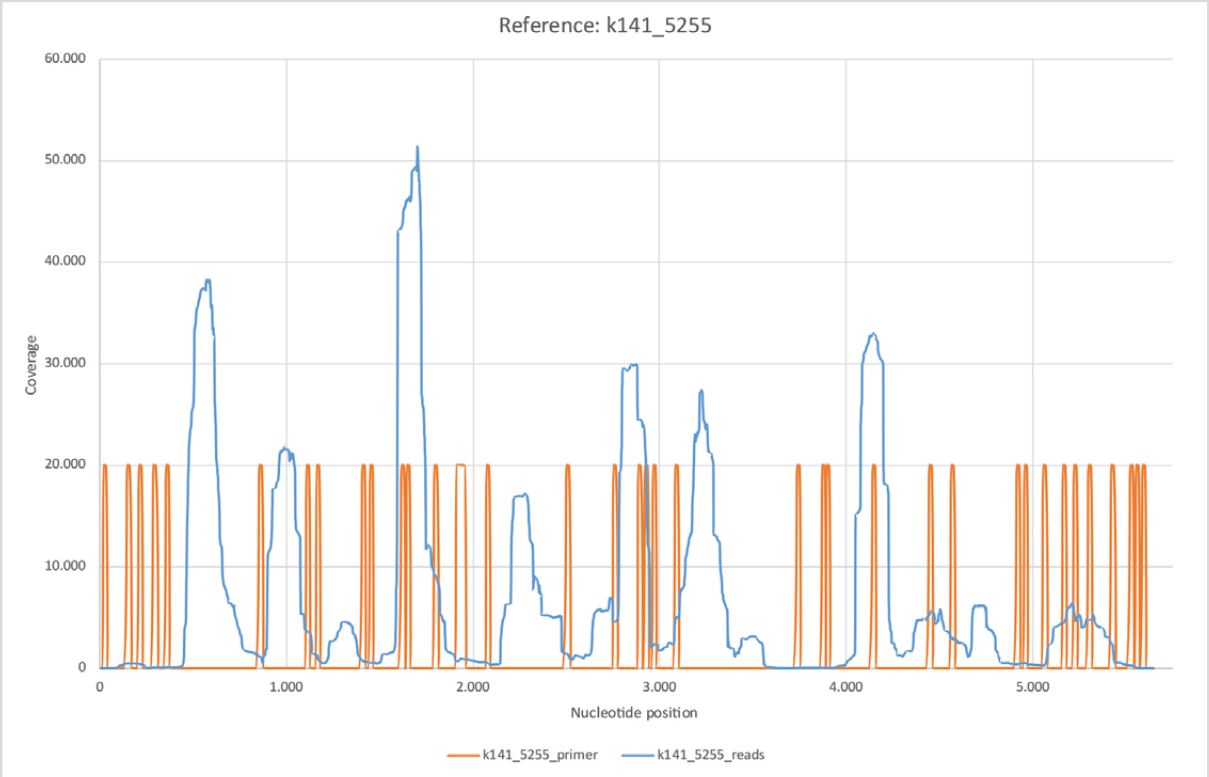
Abbildung 15: *Referenz rnd_wh_mk_1.*
a) rnd_wh_mk_1 _reads gemappt mit BBMap (M1; M2) = (37; 0,60). b) rnd_wh_mk_1_primer gemappt mit BBMap.
c) Die Abdeckungsverteilung zeigt weitestgehend eine zufällige Verteilung, vergleichbar mit Abbildung 11.



Reference - rnd_wh_mk_2	
Genome length	29.903
Number of reads	52.557
Ø Read length	43,33
P(Covering a nucleotide)	0,00144898
EN (Expected coverage)	76,1541
VARN (Binomial distribution)	76,0438
Covered nucleotides	29.903
Coverage in %	100,00%

Primer	
Genome length	29.903
Number of reads	51
Ø Read length	23,75
P(Covering a nucleotide)	0,00079407
EN (Expected coverage)	0,0405
VARN (Binomial distribution)	0,0405
Covered nucleotides	1.196
Coverage in %	4,00%
Error rate in %	34,00%

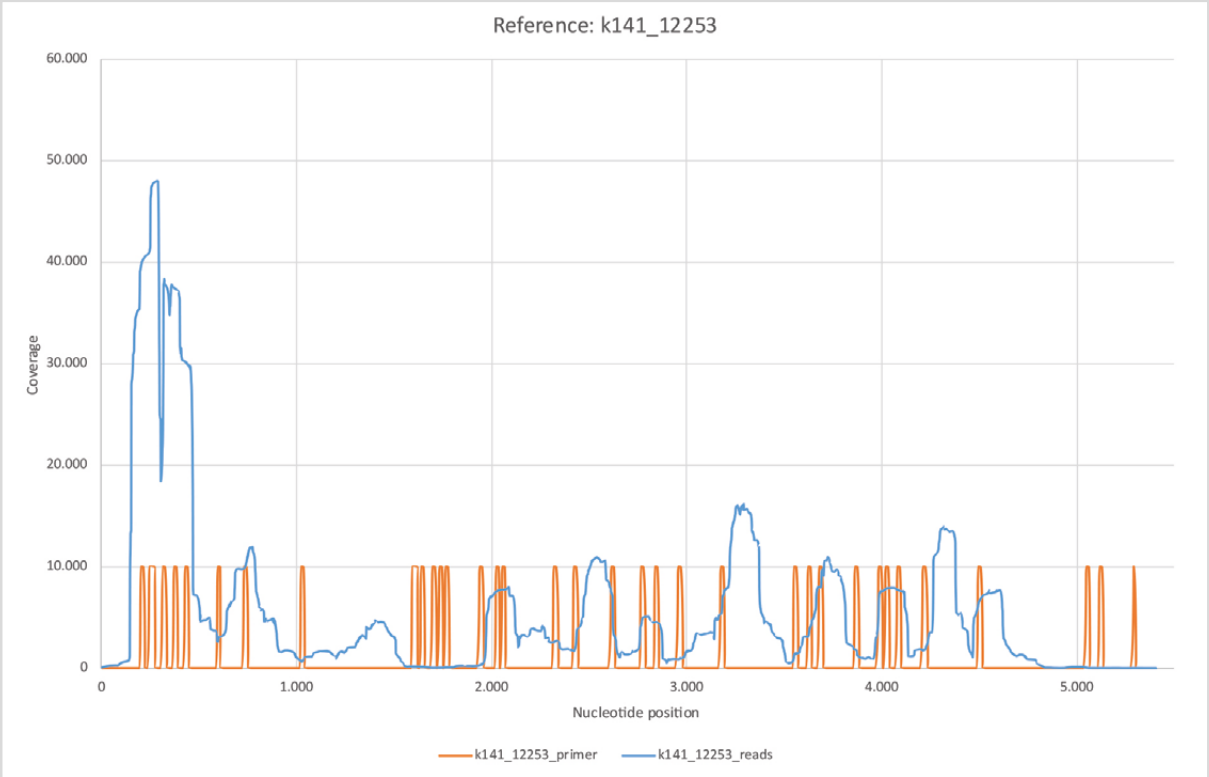
Abbildung 16: *Referenz rnd_wh_mk_2.*
a) rnd_wh_mk_2_reads gemappt mit BBMap (M1; M2) = (37; 0,60). b) rnd_wh_mk_2_primer gemappt mit BBMap.
c) Die Abdeckungsverteilung zeigt weitestgehend eine zufällige Verteilung, vergleichbar mit Abbildung 11.



Reference - k141_5255	
Genome length	5.659
Number of reads	290.188
Ø Read length	139,32
P(Covering a nucleotide)	0,02461840
EN (Expected coverage)	7143,9645
VARN (Binomial distribution)	6968,0915
Covered nucleotides	5.659
Coverage in %	100,00%

Primer	
Genome length	5.659
Number of reads	40
Ø Read length	23,13
P(Covering a nucleotide)	0,00408641
EN (Expected coverage)	0,1635
VARN (Binomial distribution)	0,1628
Covered nucleotides	895
Coverage in %	15,82%
Error rate in %	37,50%

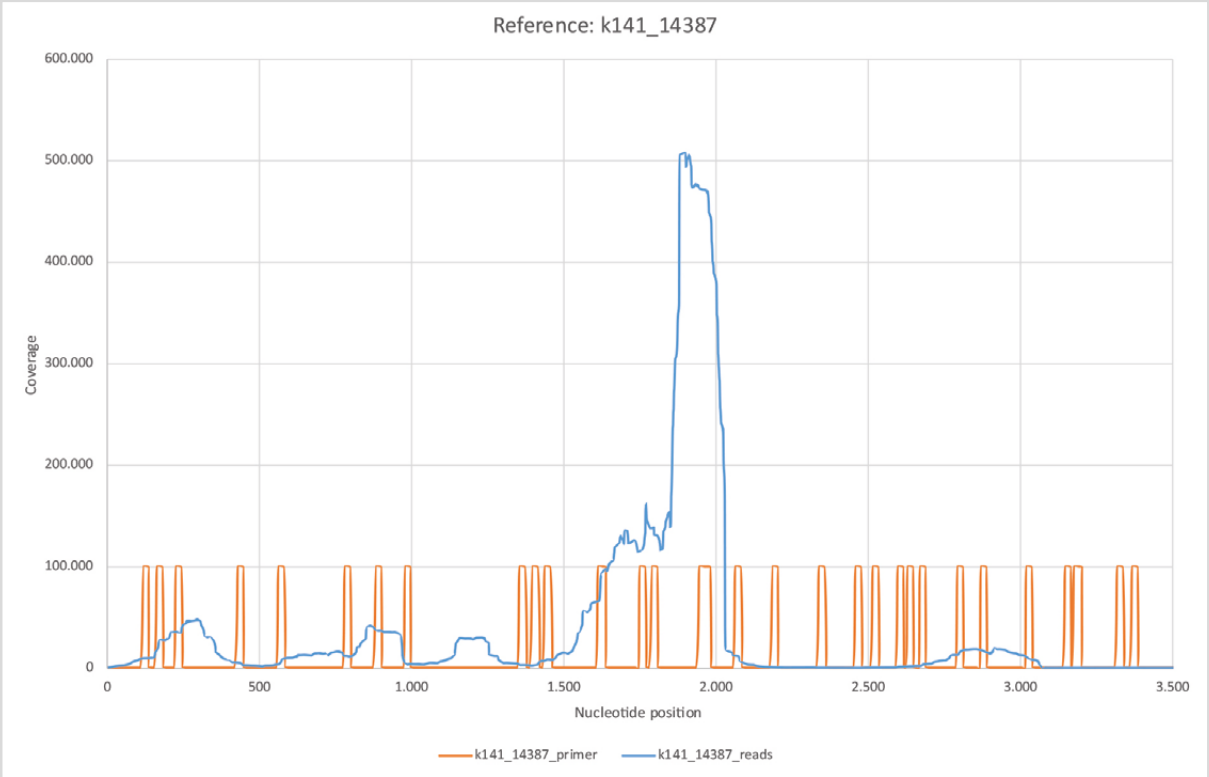
Abbildung 17: *Referenz k141_5255.*
a) k141_5255_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) k141_5255_primer gemappt mit BBMap.



Reference - k141_12253	
Genome length	5.414
Number of reads	213.744
Ø Read length	142,04
P(Covering a nucleotide)	0,02623561
EN (Expected coverage)	5607,7039
VARN (Binomial distribution)	5460,5824
Covered nucleotides	5.414
Coverage in %	100,00%

Primer	
Genome length	5.414
Number of reads	38
Ø Read length	22,82
P(Covering a nucleotide)	0,00421422
EN (Expected coverage)	0,1601
VARN (Binomial distribution)	0,1595
Covered nucleotides	812
Coverage in %	15,00%
Error rate in %	37,30%

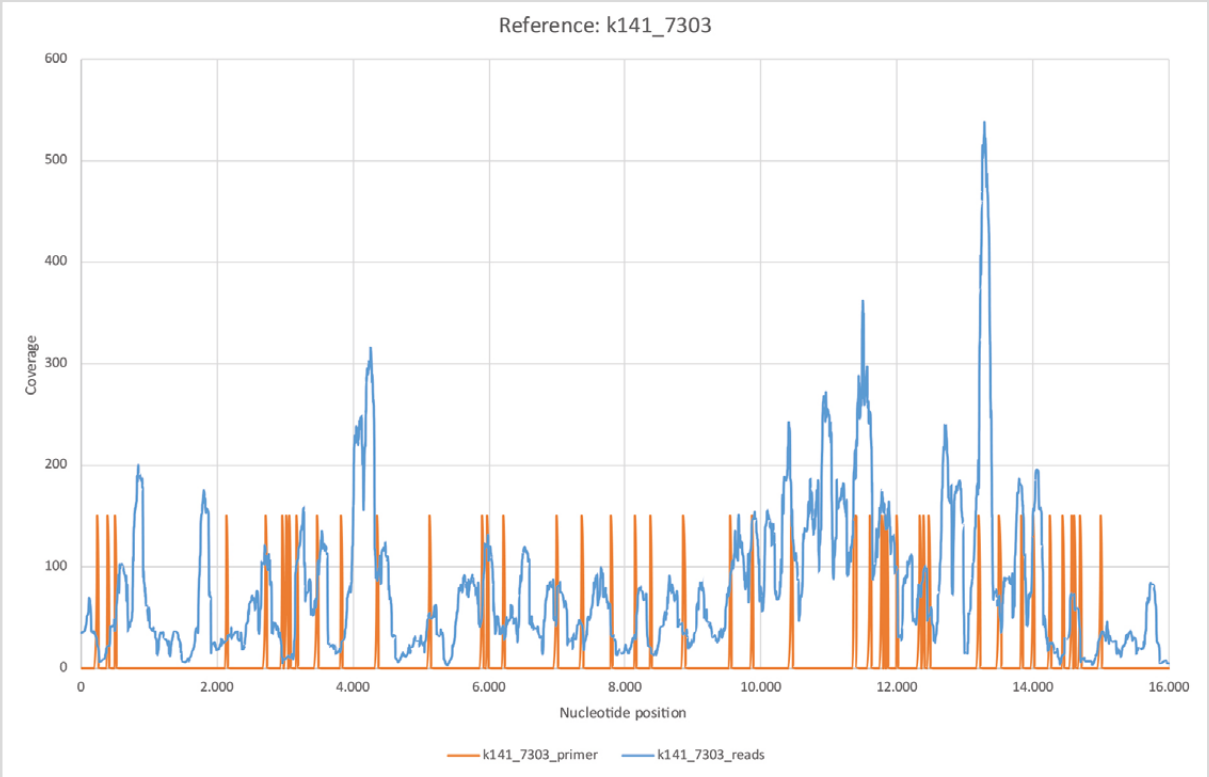
Abbildung 18: *Referenz k141_12253.*
a) k141_12253_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) k141_12253_primer gemappt mit BBMap.



Reference - k141_14387	
Genome length	3.526
Number of reads	933.303
Ø Read length	144,53
P(Covering a nucleotide)	0,04098975
EN (Expected coverage)	38255,8522
VARN (Binomial distribution)	36687,7546
Covered nucleotides	3.526
Coverage in %	100,00%

Primer	
Genome length	3.526
Number of reads	33
Ø Read length	23,06
P(Covering a nucleotide)	0,00654016
EN (Expected coverage)	0,2158
VARN (Binomial distribution)	0,2144
Covered nucleotides	722
Coverage in %	20,48%
Error rate in %	37,80%

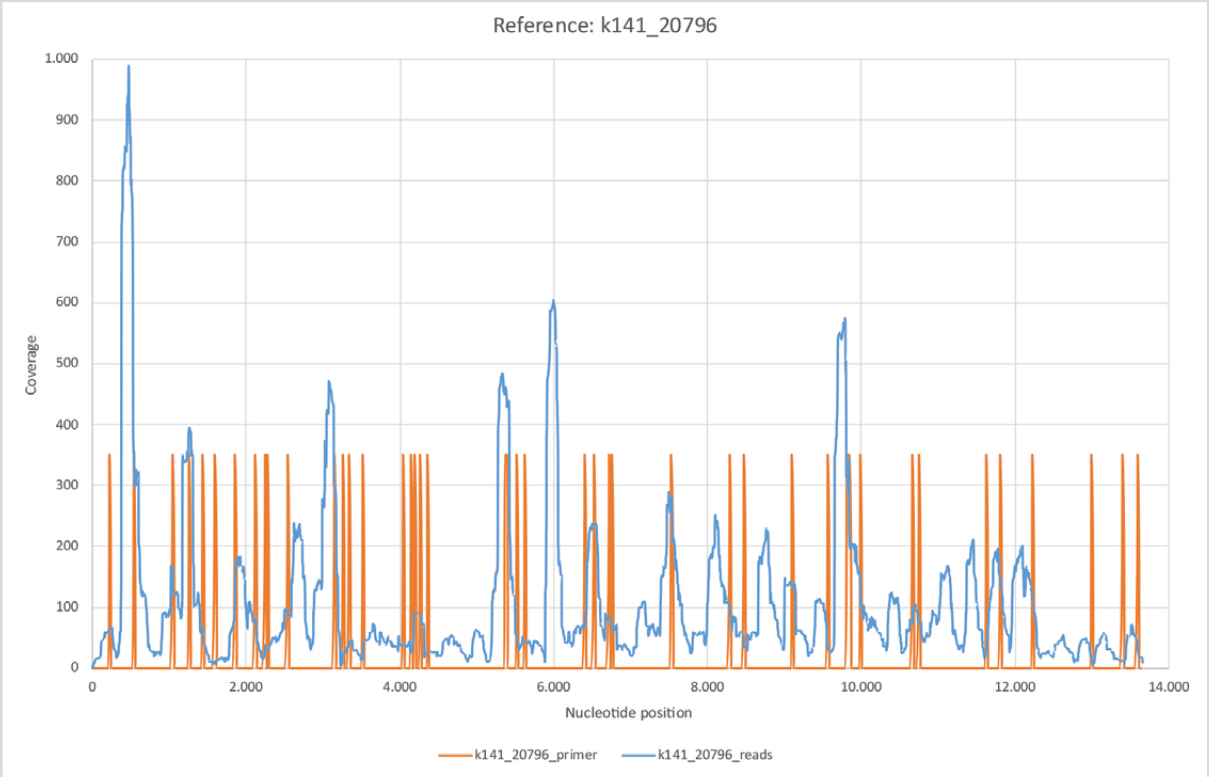
Abbildung 19: *Referenz k141_14387.*
a) k141_14387_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) k141_14387_primer gemappt mit BBMap.



Reference - k141_7303	
Genome length	16.036
Number of reads	9.138
Ø Read length	138,63
P(Covering a nucleotide)	0,00864505
EN (Expected coverage)	78,9985
VARN (Binomial distribution)	78,3156
Covered nucleotides	16.036
Coverage in %	100,00%

Primer	
Genome length	16.036
Number of reads	50
Ø Read length	23,52
P(Covering a nucleotide)	0,00146670
EN (Expected coverage)	0,0733
VARN (Binomial distribution)	0,0732
Covered nucleotides	1.091
Coverage in %	6,80%
Error rate in %	35,80%

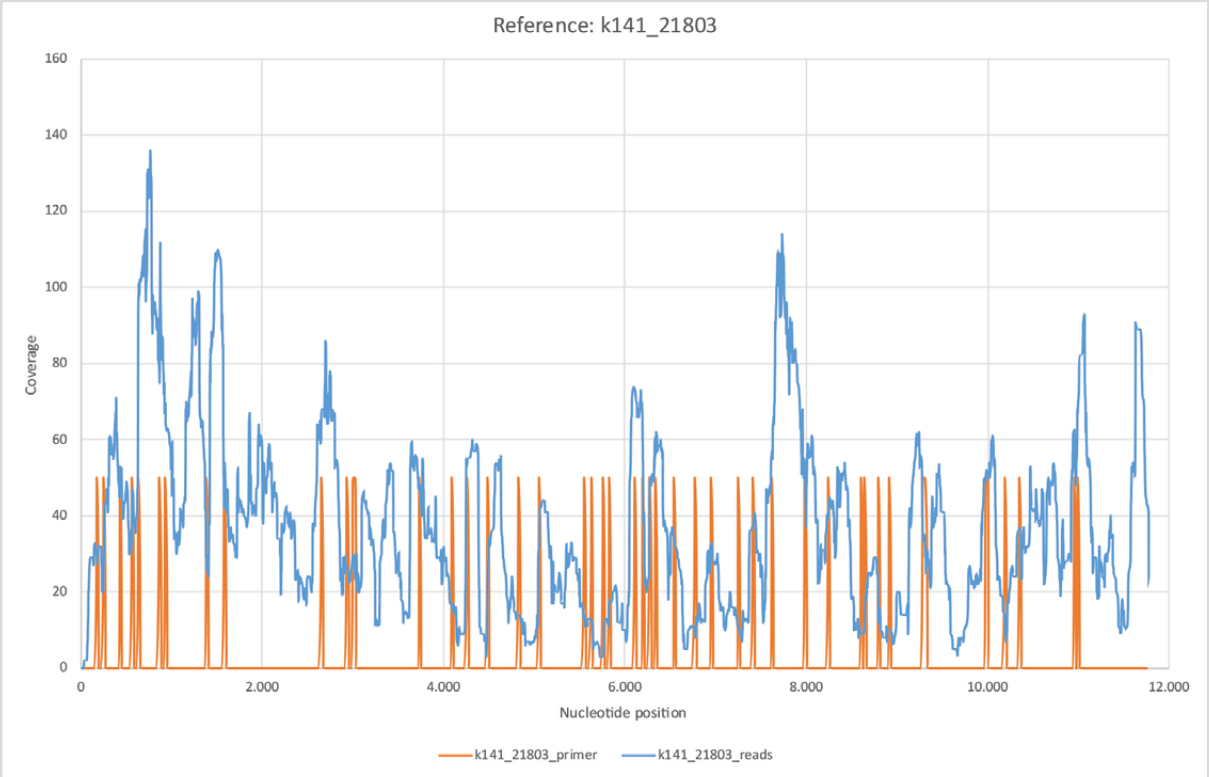
Abbildung 20: *Referenz k141_7303.*
a) k141_7303_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) k141_7303_primer gemappt mit BBMap.



Reference - k141_20796	
Genome length	13.656
Number of reads	10.287
Ø Read length	142,11
P(Covering a nucleotide)	0,01040648
EN (Expected coverage)	107,0515
VARN (Binomial distribution)	105,9374
Covered nucleotides	13.645
Coverage in %	99,92%

Primer	
Genome length	13.656
Number of reads	47
Ø Read length	23,49
P(Covering a nucleotide)	0,00172008
EN (Expected coverage)	0,0808
VARN (Binomial distribution)	0,0807
Covered nucleotides	1.053
Coverage in %	7,71%
Error rate in %	35,80%

Abbildung 21: *Referenz k141_20796.*
a) k141_20796_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) k141_20796_primer gemappt mit BBMap.



Reference - k141_21803	
Genome length	11.776
Number of reads	3.097
Ø Read length	141,83
P(Covering a nucleotide)	0,01204406
EN (Expected coverage)	37,3004
VARN (Binomial distribution)	36,8512
Covered nucleotides	11.744
Coverage in %	99,73%

Primer	
Genome length	11.776
Number of reads	49
Ø Read length	23,51
P(Covering a nucleotide)	0,00199645
EN (Expected coverage)	0,0978
VARN (Binomial distribution)	0,0976
Covered nucleotides	1.099
Coverage in %	9,33%
Error rate in %	38,50%

Abbildung 22: *Referenz k141_21803.*
a) k141_21803_reads gemappt mit Bowtie2 unter Verwendung der Standardeinstellungen. b) k141_21803_primer gemappt mit BBMap.

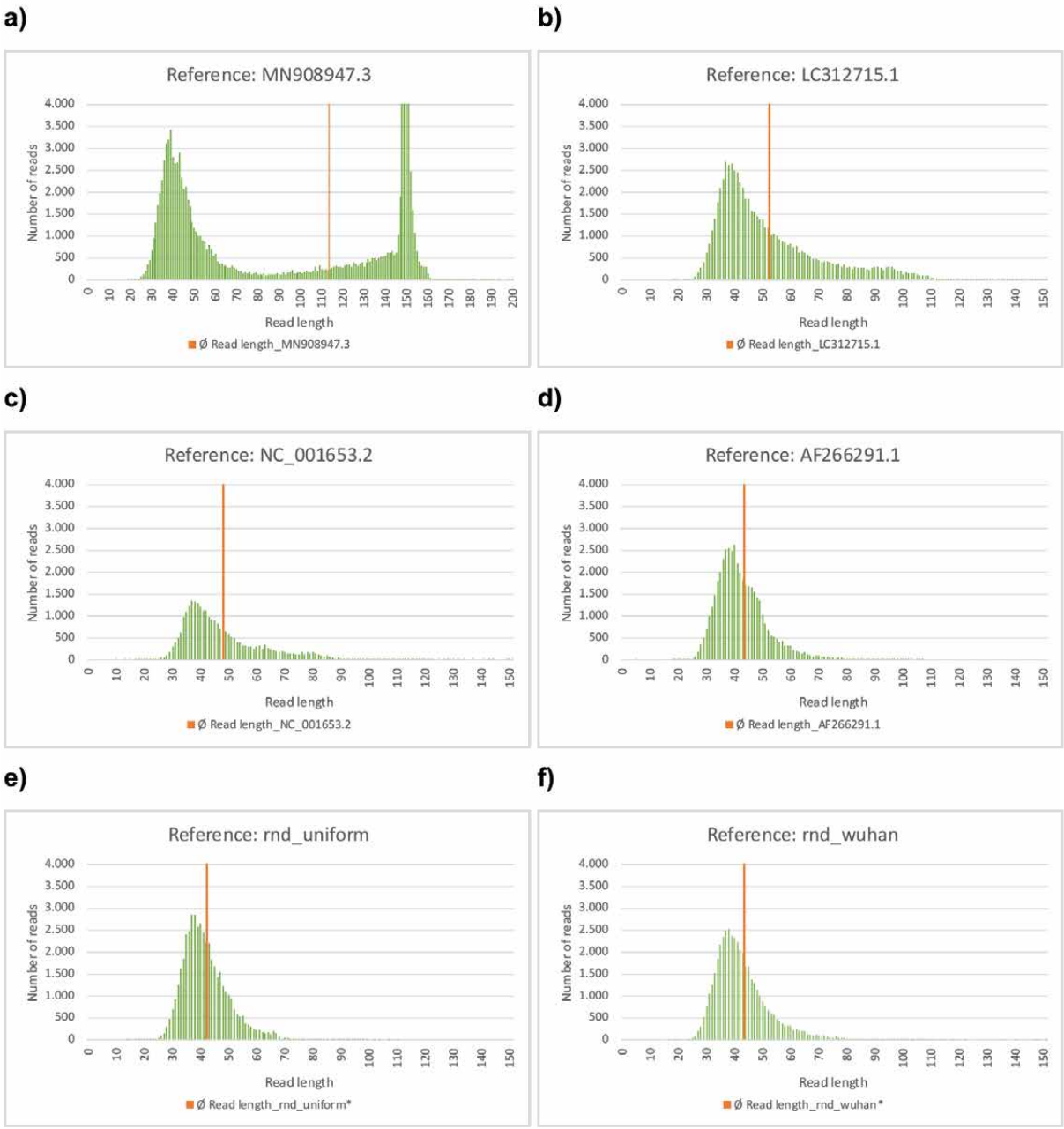


Abbildung 23:
a)–f) Mapping mit BBMap (M1; M2) = (37; 0,60). Analyse in Excel.

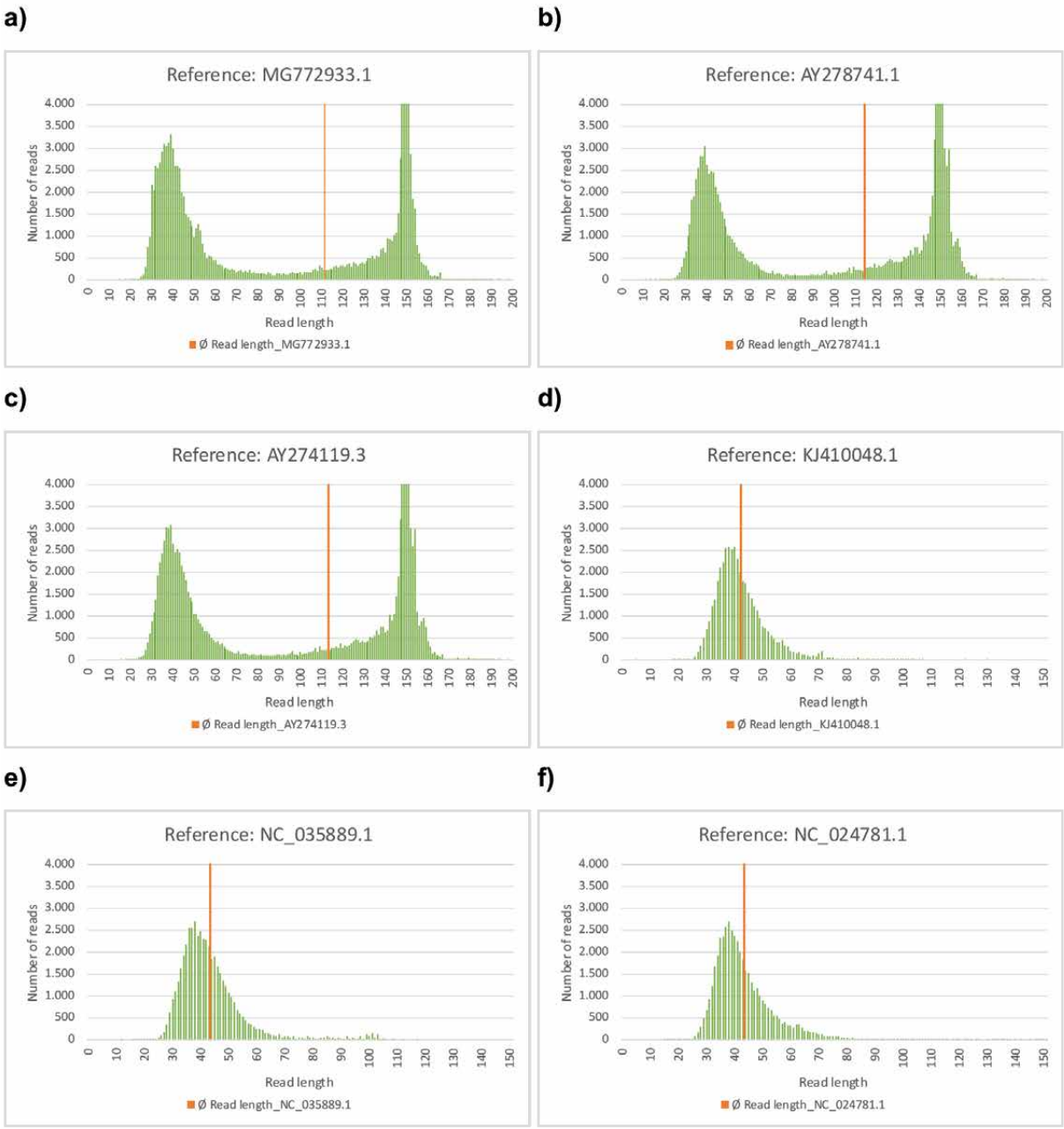


Abbildung 24:
a)–f) Mapping mit BBMap (M1; M2) = (37; 0,60). Analyse in Excel.

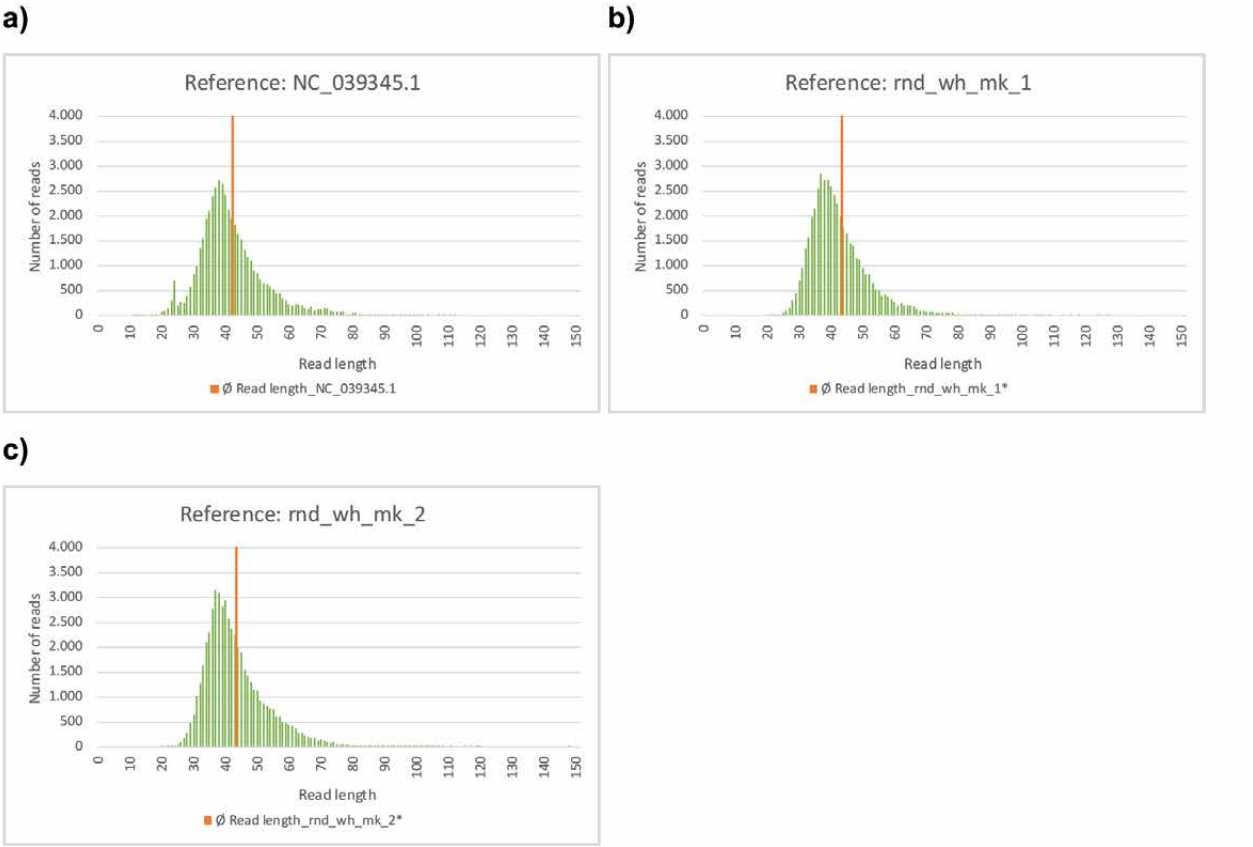
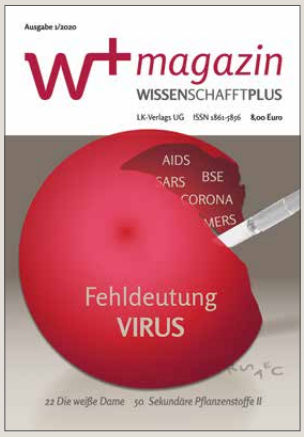
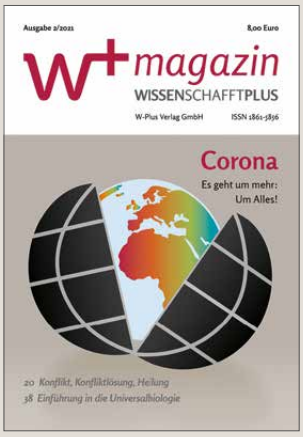


Abbildung 25:
a)–c) Mapping mit BBMap (M1; M2) = (37; 0,60). Analyse in Excel.

w+ magazin
Abonnement





Abonnieren Sie jährlich 4 Ausgaben des w+ magazins:

als gedrucktes Heft:	29 Euro
als PDF per E-Mail:	18 Euro
oder gedruckt+PDF:	38 Euro

unter www.wissenschafftplus.de

Bestellen Sie eine kostenlose Probeausgabe (als PDF oder Print) von Wissenschafftplus

per E-Mail: info@wplus-verlag.de,
Fax: 03327-5708930,
oder telefonisch: 03327 7269079