MUSTv2: an improved *de novo* detection program for recently active miniature inverted repeat transposable elements (MITEs)

Ruiquan Ge[1,*], Guoqin Mai[2,*], Ruochi Zhang[3,*], Qing Wu[1], Xundong Wu[1], Fengfeng Zhou[3,#].

1. School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, Zhejiang, China, 310018.
2. Center for Synthetic Biology Engineering Research, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China, 518055.
3. College of Computer Science and Technology, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin, China, 130012.

[*] These authors contribute equally to this work.
[#] Corresponding author: Fengfeng Zhou, phone: +86-431-85166024; e-mail: FengfengZhou@gmail.com, or ffzhou@jlu.edu.cn. Web site: http://www.healthinformaticslab.org/ffzhou/ .

Supplementary Table S1

50 copies of the three known MITEs, *i.e.* Chunjie, mPing, and Nezha, are randomly inserted into the chromosome of E. coli K12 MG1655 (chromosome ID: NC_000913.2). Column "Pos" is the actual position in the chromosome NC_000913.2 where the simulated insertion occurs. Since the major features of a MITE are TIRs and DRs, where exist no matter which strand it is on, all the insertions occur in the positive strand.

#Chr	RawID	MITE-name	Start	End	Pos	MITELength	TSDLength
EcoliK12MG1655	Chunjie|2	Chunjie	18272	18490	18262	219	9
EcoliK12MG1655	Chunjie|25	Chunjie	42686	42904	42439	219	9
EcoliK12MG1655	Nezha|45	Nezha	55097	55234	54612	138	10
EcoliK12MG1655	mPing|35	mPing	108904	109333	108268	430	3
EcoliK12MG1655	mPing|17	mPing	160931	161360	159859	430	3
EcoliK12MG1655	Chunjie|26	Chunjie	254893	255111	253379	219	9
EcoliK12MG1655	Nezha|3	Nezha	300431	300568	298679	138	10
EcoliK12MG1655	mPing|30	mPing	307243	307672	305340	430	3
EcoliK12MG1655	Nezha|10	Nezha	361883	362020	359537	138	10
EcoliK12MG1655	mPing|33	mPing	384198	384627	381701	430	3

EcoliK12MG1655	Chunjie|44	Chunjie	397883	398101	394944	219	9
EcoliK12MG1655	Chunjie|39	Chunjie	422495	422713	419319	219	9
EcoliK12MG1655	Chunjie|30	Chunjie	426968	427186	423555	219	9
EcoliK12MG1655	mPing|28	mPing	428432	428861	424788	430	3
EcoliK12MG1655	Nezha|37	Nezha	430670	430807	426583	138	10
EcoliK12MG1655	mPing|27	mPing	434315	434744	430077	430	3
EcoliK12MG1655	Nezha|0	Nezha	489544	489681	484863	138	10
EcoliK12MG1655	Chunjie|47	Chunjie	509805	510023	504967	219	9
EcoliK12MG1655	Chunjie|11	Chunjie	518517	518735	513442	219	9
EcoliK12MG1655	mPing|8	mPing	542180	542609	536874	430	3
EcoliK12MG1655	mPing|25	mPing	549184	549613	543442	430	3
EcoliK12MG1655	Nezha|31	Nezha	590647	590784	584462	138	10
EcoliK12MG1655	Chunjie|0	Chunjie	643121	643339	636779	219	9
EcoliK12MG1655	Nezha|20	Nezha	692438	692575	685858	138	10
EcoliK12MG1655	Chunjie|32	Chunjie	811482	811700	804745	219	9
EcoliK12MG1655	Nezha|27	Nezha	874999	875136	868024	138	10
EcoliK12MG1655	Nezha|32	Nezha	928474	928611	921341	138	10
EcoliK12MG1655	Nezha|26	Nezha	928810	928947	921519	138	10
EcoliK12MG1655	Chunjie|33	Chunjie	949818	950036	942370	219	9
EcoliK12MG1655	Nezha|22	Nezha	961853	961990	954167	138	10
EcoliK12MG1655	Nezha|24	Nezha	966817	966954	958973	138	10
EcoliK12MG1655	Nezha|28	Nezha	971695	971832	963693	138	10
EcoliK12MG1655	Chunjie|28	Chunjie	1003876	1004094	995717	219	9
EcoliK12MG1655	Chunjie|45	Chunjie	1074240	1074458	1065844	219	9
EcoliK12MG1655	mPing|12	mPing	1075777	1076206	1067150	430	3
EcoliK12MG1655	Chunjie|22	Chunjie	1104157	1104375	1095088	219	9
EcoliK12MG1655	Nezha|46	Nezha	1145617	1145754	1136310	138	10
EcoliK12MG1655	mPing|41	mPing	1162829	1163258	1153371	430	3
EcoliK12MG1655	Nezha|42	Nezha	1169080	1169217	1159179	138	10
EcoliK12MG1655	Nezha|47	Nezha	1233079	1233216	1223020	138	10
EcoliK12MG1655	mPing|26	mPing	1329690	1330119	1319480	430	3
EcoliK12MG1655	mPing|5	mPing	1353466	1353895	1342820	430	3
EcoliK12MG1655	mPing|37	mPing	1415225	1415654	1404143	430	3
EcoliK12MG1655	mPing|21	mPing	1431342	1431771	1419824	430	3
EcoliK12MG1655	Chunjie|34	Chunjie	1460550	1460768	1448590	219	9
EcoliK12MG1655	Nezha|21	Nezha	1493937	1494074	1481739	138	10
EcoliK12MG1655	Nezha|1	Nezha	1523994	1524131	1511638	138	10
EcoliK12MG1655	Chunjie|49	Chunjie	1534958	1535176	1522445	219	9
EcoliK12MG1655	Chunjie|7	Chunjie	1600023	1600241	1587273	219	9
EcoliK12MG1655	Nezha|13	Nezha	1642406	1642543	1629418	138	10
EcoliK12MG1655	Chunjie|29	Chunjie	1739375	1739593	1726230	219	9
EcoliK12MG1655	Nezha|30	Nezha	1754926	1755063	1741543	138	10
EcoliK12MG1655	Nezha|23	Nezha	1803313	1803450	1789772	138	10

EcoliK12MG1655	Chunjie|1	Chunjie	1807515	1807733	1793817	219	9
EcoliK12MG1655	mPing|9	mPing	1815803	1816232	1801874	430	3
EcoliK12MG1655	Nezha|33	Nezha	1941266	1941403	1926894	138	10
EcoliK12MG1655	mPing|18	mPing	1963695	1964124	1949172	430	3
EcoliK12MG1655	mPing|10	mPing	2040146	2040575	2025187	430	3
EcoliK12MG1655	Chunjie|18	Chunjie	2071907	2072125	2056506	219	9
EcoliK12MG1655	mPing|19	mPing	2074593	2075022	2058961	430	3
EcoliK12MG1655	Chunjie|46	Chunjie	2088790	2089008	2072716	219	9
EcoliK12MG1655	Nezha|25	Nezha	2122456	2122593	2106144	138	10
EcoliK12MG1655	Nezha|2	Nezha	2180548	2180685	2164078	138	10
EcoliK12MG1655	mPing|47	mPing	2183567	2183996	2166946	430	3
EcoliK12MG1655	mPing|49	mPing	2215889	2216318	2198832	430	3
EcoliK12MG1655	Nezha|8	Nezha	2238575	2238712	2221075	138	10
EcoliK12MG1655	Chunjie|24	Chunjie	2250341	2250559	2232684	219	9
EcoliK12MG1655	Chunjie|4	Chunjie	2257885	2258103	2239991	219	9
EcoliK12MG1655	Nezha|39	Nezha	2261655	2261792	2243523	138	10
EcoliK12MG1655	mPing|34	mPing	2273958	2274387	2255675	430	3
EcoliK12MG1655	Chunjie|35	Chunjie	2284175	2284393	2265450	219	9
EcoliK12MG1655	Nezha|43	Nezha	2323366	2323503	2304403	138	10
EcoliK12MG1655	mPing|15	mPing	2331834	2332263	2312720	430	3
EcoliK12MG1655	Nezha|4	Nezha	2348661	2348798	2329104	138	10
EcoliK12MG1655	Chunjie|42	Chunjie	2350397	2350615	2330683	219	9
EcoliK12MG1655	Nezha|36	Nezha	2366069	2366206	2346117	138	10
EcoliK12MG1655	mPing|1	mPing	2373435	2373864	2353332	430	3
EcoliK12MG1655	Chunjie|5	Chunjie	2403205	2403423	2382660	219	9
EcoliK12MG1655	Nezha|48	Nezha	2444601	2444738	2423818	138	10
EcoliK12MG1655	Nezha|41	Nezha	2498280	2498417	2477339	138	10
EcoliK12MG1655	Chunjie|19	Chunjie	2539300	2539518	2518202	219	9
EcoliK12MG1655	mPing|45	mPing	2623072	2623501	2601743	430	3
EcoliK12MG1655	mPing|2	mPing	2645422	2645851	2623657	430	3
EcoliK12MG1655	Nezha|18	Nezha	2676775	2676912	2654567	138	10
EcoliK12MG1655	Nezha|19	Nezha	2701793	2701930	2679427	138	10
EcoliK12MG1655	Nezha|15	Nezha	2722016	2722153	2699492	138	10
EcoliK12MG1655	Nezha|12	Nezha	2760732	2760869	2738050	138	10
EcoliK12MG1655	Chunjie|3	Chunjie	2781736	2781954	2758897	219	9
EcoliK12MG1655	Nezha|38	Nezha	2796813	2796950	2773736	138	10
EcoliK12MG1655	Chunjie|27	Chunjie	2847190	2847408	2823956	219	9
EcoliK12MG1655	Nezha|14	Nezha	2880410	2880547	2856938	138	10
EcoliK12MG1655	Chunjie|48	Chunjie	2910214	2910432	2886585	219	9
EcoliK12MG1655	mPing|7	mPing	2922741	2923170	2898881	430	3
EcoliK12MG1655	mPing|31	mPing	2953731	2954160	2929435	430	3
EcoliK12MG1655	mPing|24	mPing	2966877	2967306	2942145	430	3
EcoliK12MG1655	Nezha|44	Nezha	2979041	2979178	2953866	138	10

EcoliK12MG1655	Nezha|9	Nezha	3013896	3014033	2988563	138	10
EcoliK12MG1655	mPing|0	mPing	3016044	3016473	2990560	430	3
EcoliK12MG1655	mPing|48	mPing	3017895	3018324	2991975	430	3
EcoliK12MG1655	mPing|39	mPing	3021369	3021798	2995013	430	3
EcoliK12MG1655	Chunjie|43	Chunjie	3028346	3028564	3001548	219	9
EcoliK12MG1655	Chunjie|6	Chunjie	3067678	3067896	3040643	219	9
EcoliK12MG1655	mPing|46	mPing	3109055	3109484	3081789	430	3
EcoliK12MG1655	mPing|44	mPing	3137672	3138101	3109970	430	3
EcoliK12MG1655	mPing|3	mPing	3189691	3190120	3161553	430	3
EcoliK12MG1655	Chunjie|10	Chunjie	3217667	3217885	3189087	219	9
EcoliK12MG1655	Nezha|49	Nezha	3233253	3233390	3204435	138	10
EcoliK12MG1655	Chunjie|13	Chunjie	3238599	3238817	3209624	219	9
EcoliK12MG1655	Chunjie|38	Chunjie	3277873	3278091	3248661	219	9
EcoliK12MG1655	Chunjie|17	Chunjie	3287371	3287589	3257922	219	9
EcoliK12MG1655	mPing|6	mPing	3294528	3294957	3264848	430	3
EcoliK12MG1655	Nezha|5	Nezha	3308503	3308640	3278380	138	10
EcoliK12MG1655	Chunjie|40	Chunjie	3311945	3312163	3281665	219	9
EcoliK12MG1655	Chunjie|9	Chunjie	3345673	3345891	3315156	219	9
EcoliK12MG1655	mPing|4	mPing	3391410	3391839	3360662	430	3
EcoliK12MG1655	Nezha|34	Nezha	3404186	3404323	3372995	138	10
EcoliK12MG1655	mPing|20	mPing	3420950	3421379	3389608	430	3
EcoliK12MG1655	Nezha|40	Nezha	3442580	3442717	3410795	138	10
EcoliK12MG1655	Nezha|6	Nezha	3499176	3499313	3467233	138	10
EcoliK12MG1655	Chunjie|36	Chunjie	3535159	3535377	3503059	219	9
EcoliK12MG1655	Nezha|11	Nezha	3542390	3542527	3510052	138	10
EcoliK12MG1655	Chunjie|16	Chunjie	3567690	3567908	3535195	219	9
EcoliK12MG1655	mPing|16	mPing	3614137	3614566	3581411	430	3
EcoliK12MG1655	Chunjie|31	Chunjie	3639880	3640098	3606712	219	9
EcoliK12MG1655	mPing|11	mPing	3642088	3642517	3608689	430	3
EcoliK12MG1655	Chunjie|20	Chunjie	3679206	3679424	3645365	219	9
EcoliK12MG1655	Chunjie|15	Chunjie	3703834	3704052	3669756	219	9
EcoliK12MG1655	mPing|32	mPing	3843203	3843632	3808894	430	3
EcoliK12MG1655	mPing|43	mPing	3855627	3856056	3820882	430	3
EcoliK12MG1655	mPing|42	mPing	3928201	3928630	3893020	430	3
EcoliK12MG1655	Chunjie|14	Chunjie	3964376	3964594	3928753	219	9
EcoliK12MG1655	Chunjie|23	Chunjie	3986927	3987145	3951067	219	9
EcoliK12MG1655	mPing|14	mPing	4043939	4044368	4007848	430	3
EcoliK12MG1655	Chunjie|41	Chunjie	4071141	4071359	4034608	219	9
EcoliK12MG1655	Nezha|29	Nezha	4086305	4086442	4049534	138	10
EcoliK12MG1655	mPing|23	mPing	4125071	4125500	4088149	430	3
EcoliK12MG1655	Chunjie|8	Chunjie	4151125	4151343	4113761	219	9
EcoliK12MG1655	mPing|22	mPing	4228403	4228832	4190808	430	3
EcoliK12MG1655	mPing|36	mPing	4229387	4229816	4191356	430	3

EcoliK12MG1655	mPing|40	mPing	4265590	4266019	4227123	430	3
EcoliK12MG1655	Chunjie|21	Chunjie	4295878	4296096	4256969	219	9
EcoliK12MG1655	mPing|29	mPing	4315277	4315706	4276137	430	3
EcoliK12MG1655	Nezha|7	Nezha	4316532	4316669	4276949	138	10
EcoliK12MG1655	Chunjie|37	Chunjie	4344914	4345132	4305174	219	9
EcoliK12MG1655	Chunjie|12	Chunjie	4367849	4368067	4327872	219	9
EcoliK12MG1655	Nezha|35	Nezha	4409275	4409412	4369060	138	10
EcoliK12MG1655	mPing|38	mPing	4514864	4515293	4474498	430	3
EcoliK12MG1655	Nezha|17	Nezha	4537710	4537847	4496901	138	10
EcoliK12MG1655	Nezha|16	Nezha	4563971	4564108	4523004	138	10
EcoliK12MG1655	mPing|13	mPing	4568669	4569098	4527551	430	3

Supplementary Table S2

There are 136 DNA miniature elements with at least 10 full copies recovered in the rice genome. The full copy detection sensitivities of the two tools MITE-Hunter and MUSTv2 were calculated in the following table. The minimum full copy number 10 is facilitated for the statistical meaningfulness of the calculated sensitivity, and specificity is not calculated, due to that there is no commonly recognized negative dataset. Column "Repeat" is the name of the repeat, "Group" is the repeat group name. The columns "RepeatMasker", "MITE-Hunter" and "MUSTv2" are the full copy numbers of this repeat in the rice genome detected by the tool RepeatMasker, MITE-Hunter and MUSTv2, respectively. The percentile distributions of the two tools MITE-Hunter and MUSTv2 may be found in the Supplementary Figure S2, after this table.

#Repeat	Group	RepeatMasker	MITE-Hunter	MUSTv2
AMYLTP		211	209	20
BUHUI	Tourist	162	152	13
CASGRANDA	Tourist	16	0	2
CASLET	Tourist	30	8	9
CASTAWAY	Tourist	824	575	204
CASTAWAY-3	Tourist	29	26	2
CENTRE	Tourist	11	3	1
CLOUD-3	MuDR	97	0	9
CLOUD-4	MuDR	146	0	13
CLOUD-6	MuDR	63	0	3
CLOUD-7	MuDR	69	0	1
COWARD	Tourist	206	144	20
COWARD-3	Tourist	564	456	118
DEBOAT	hAT	72	71	25
DELAY	hAT	57	55	8
DITAILA	Tourist	95	0	5
DITTO	Tourist	1702	61	181
DITTO-2	Tourist	187	2	22

DITTO3	Tourist	393	37	29
DS-RICE3N	hAT-Ac	57	0	5
ECR	MuDR	103	101	30
ECSR	Tourist	31	0	3
Explorer		245	0	17
F118	hAT	157	155	21
F1275	Tourist	37	0	3
F569	Tourist	273	221	47
F770	Tourist	180	41	54
F804		115	0	6
FOCUS	Tourist	70	39	4
Gaijin	Tourist	2165	0	218
GLUTEL1LIKE	Tourist	19	0	1
HELIA	Tourist	10	2	1
ID-2	Tourist	13	0	0
ID-3	Tourist	147	145	14
ID-4	Tourist	103	0	10
INDITTO	Tourist	505	500	35
JINHUA	hAT-Ac	89	77	4
JOUZHEN	hAT-Ac	17	0	2
KIDDOA	Tourist	44	21	0
KIDDOB	Tourist	15	12	2
KIDDOC	Tourist	15	15	5
KIDDOD	Tourist	29	24	2
MDM1	MuDR	85	67	7
MDM2	MuDR	106	0	4
MERMITE18A	MuDR	24	0	1
MPING	Harbinger	51	51	48
MUDRN4_OS	MuDR?	23	0	0
MUDRN5_OS	MuDR	40	40	10
NDNA2TNA_OS	Tourist	140	2	16
NONAME	Tourist	11	0	2
OLO24	Tourist	440	327	55
OLO24B	Tourist	211	35	21
OLO24C	Tourist	484	0	26
OSTE18		67	0	4
OSTE19	MuDR	63	0	4
OSTE20	MuDR	54	0	6
OSTE23		78	0	7
OSTE28	MuDR	55	53	2
OSTE30	hAT?	81	77	12
OSTE9		49	0	8
QINNIU	hAT-Tip100	13	1	5

QIQI	Tourist	46	43	2
SNAP-OL3		97	0	11
STOLA	Tourist	134	134	35
STOLAB	Tourist	88	0	7
STOWAWAY1_OS	TcMar-Stowaway	2757	106	573
STOWAWAY10_OS	TcMar-Stowaway	350	330	71
STOWAWAY11_OS	TcMar-Stowaway	461	0	158
STOWAWAY12_OS	TcMar-Stowaway	249	0	87
STOWAWAY13_OS	TcMar-Stowaway	369	364	68
STOWAWAY14_OS	TcMar-Stowaway	35	0	8
STOWAWAY15_OS	TcMar-Stowaway	35	10	3
STOWAWAY15-2_OS	TcMar-Stowaway	134	53	13
STOWAWAY16_OS	TcMar-Stowaway	19	16	5
STOWAWAY17_OS	TcMar-Stowaway	598	172	141
STOWAWAY18_OS	TcMar-Stowaway	180	123	55
STOWAWAY2_OS	TcMar-Stowaway	1181	34	338
STOWAWAY20_OS	TcMar-Stowaway	55	55	10
STOWAWAY21_OS	TcMar-Stowaway	1253	0	207
STOWAWAY23_OS	TcMar-Stowaway	15	3	0
STOWAWAY24_OS	TcMar-Stowaway	11	0	2
STOWAWAY25_OS	TcMar-Stowaway	52	28	6
STOWAWAY26_OS	TcMar-Stowaway	22	3	2
STOWAWAY29_OS	TcMar-Stowaway	28	0	2
STOWAWAY30-2_OS	TcMar-Stowaway	10	4	2
STOWAWAY31_OS	TcMar-Stowaway	15	8	0
STOWAWAY32_OS	TcMar-Stowaway	71	65	8
STOWAWAY34_OS	TcMar-Stowaway	132	130	16
STOWAWAY35_OS	TcMar-Stowaway	80	0	7
STOWAWAY36_OS	TcMar-Stowaway	134	130	30
STOWAWAY40_OS	TcMar-Stowaway	15	0	1
STOWAWAY41_OS	TcMar-Stowaway	1745	0	140
STOWAWAY42_OS	TcMar-Stowaway	201	0	14
STOWAWAY43_OS	TcMar-Stowaway	563	0	55
STOWAWAY44_OS	TcMar-Stowaway	803	0	56
STOWAWAY45_OS	TcMar-Stowaway	228	0	20
STOWAWAY46_OS	TcMar-Stowaway	505	0	39
STOWAWAY47_OS	TcMar-Stowaway	1253	0	99
STOWAWAY48_OS	TcMar-Stowaway	196	0	15
STOWAWAY49_OS	TcMar-Stowaway	75	0	8
STOWAWAY50_OS	TcMar-Stowaway	499	0	36
STOWAWAY51_OS	TcMar-Stowaway	290	0	22
STOWAWAY8_OS	TcMar-Stowaway	475	467	42
STOWAWAY9_OS	TcMar-Stowaway	1023	0	287

SUSU	Tourist	191	0	96
TAMI2		110	0	7
TELIA		144	144	43
TEOS1		64	25	1
TESS	hAT	173	171	20
TNR12	MuDR	65	2	7
TNR2A		250	0	21
TNR8-21	MuDR	130	117	16
TNR9	hAT-Tip100	38	35	3
TOURIST6_OS	Tourist	575	0	144
TOURIST-XI	Tourist	169	152	28
TOURIST-XV	Tourist	38	0	1
TRC1	Tourist	15	14	0
TREP215	TcMar-Stowaway	2687	13	655
TREP219	TcMar-Stowaway	300	0	25
TREP220	TcMar-Stowaway	173	0	14
TYPEG	MuDR?	90	86	12
TYPEH		293	0	85
TYPEM		573	0	203
TYPEU		93	0	6
TYPEU1		94	0	3
TYPEU2		41	0	1
TYPEU3		11	0	0
TYPEU4		139	0	8
UNIQUE		464	0	54
WANDERER_OS	Tourist	633	0	79
WUJI	Tourist	20	0	1
WUWU	MuDR	29	0	0
XIKE	MuDR	28	2	4
YOUREN	Tourist	286	286	81
ZM13-1_OS	Tourist	80	0	4
ZM13-2_OS	Tourist	60	0	6