Divergence of pigments in three phylogenetically close *Monascus* species (*M. pilosus*, *M. ruber*, and *M. purpureus*) based on secondary metabolite biosynthetic gene clusters

Yuki Higa (1), Young-Soo Kim (1), Shigehiko Kanaya (2), Altaf-Ul-Amin (2), Mei Ko (2), Naoaki Ono (3)*

(1) R&D Center, Kobayashi Pharmaceutical Co., Ltd, Ibaraki-shi, Toyokawa, 1-30-3, Osaka, Japan.

(2) Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma-shi, Takayama-cho, 8916-5 Nara, Japan

(3) Data Science Center (DSC), Nara Institute of Science and Technology, Ikoma-shi, Takayama-cho, 8916-5 Nara, Japan

* Correspondence: nono@is.naist.jp

Table 1 Classification of gene clusters (ID1-54) corresponding to the dendrogram in Figure 4. The type of GCs was determined using antiSMASH software. Pi, Ru, and Pu represent the cluster ID in Figure 4. Identical gene organization is denoted by red numbers. The secondary metabolic pathways represent the secondary metabolite information based on DNA sequence homology. Type of ATP-binding cassettes detected in individual groups is represented as ATP-binding cassettes. Gene-cluster groups with both identical gene organization and high DNA sequence similarity are denoted by '**' and groups with only high DNA sequence similarity are denoted by '*'. Two-letter abbreviations use used for the Venn diagrams: Pi, *M. pilosus* NBRC 4520; Ru, *M. ruber* NBRC 4483; Pu, *M. purpureus* NBRC 4478.

ID	Type of GCs	Pi	Ru	Pu	Secondary metabolic pathways detected in DNA sequence homology	ATP-binding cassettes	Identity level
1	t1pks	1	4	-	Narbonolide/10-deoxymethynolide (*pikAI – AIV*), Phthiocerol/phenolphthionocerol (*ppsA - E*)	-	**
2	t1pks	2	-	-	Phthiocerol/phenolphthionocerol (*ppsA - E*)	-	-
3	t1pks	18	19	7	Azaphilone*, Lovastatin (LOVB, LOVF) , Narbonolide/10-deoxymethynolide (*pikAI – AIV*), Phthiocerol/phenolphthionocerol (*ppsA - E*)	-	-
4	t1pks	22	-	-	-	-	*
5	t1pks	23	48	17	Lovastatin (LOVB, LOVF) , Narbonolide/10-deoxymethynolide (*pikAI – AIV*), Phthiocerol/phenolphthionocerol (*ppsA - E*)	-	*
6	t1pks	19	43	-	Monacolin K*, Compactin, Lovastatin (LOVA, LOVB)	-	*
7	t1pks	-	-	8	Citrinin*, Narbonolide/10-deoxymethynolide (*pikAI–AIV*), Phthiocerol/phenolphthionocerol (*ppsA-C, E*)	-	*
8	t1pks	9	-	10	Byssochlamic acid*, Narbonolide/10-deoxymethynolide (*pikAI–AIV*), Phthiocerol/phenolphthionocerol (*ppsA-C, E*)	-	*
9	nrps	4	10	19	Neosartoricin*, Fengycin (*fenA–E*), Surfacin(*srfAA-AC*)	-	**
10	nrps	13	30	20	Brevianamide F (*FTMA*)	ABCB	**
11	nrps	7	2	-	Fengycin (*fenA-E*), Surfacin (srfAA-AC)	ABCC	**
12	nrps	11	23	2	Fengycin (*fenA-D*)	ABCC	**
13	nrps	5	22	12	Fengycin (*fenA-D*)	-	*
14	nrps	24	21	-	Fengycin (*ppsA-D*), surfacing (*srfAA-AC*)	ABCB	*
15	nrps	3	41	-	Fengycin (*fenA–D*)	-	*
16	nrps	-	-	9	Fengycin (*fenB, D*), Ferricrocin (*SIDC, SIDD*)	-	-
17	t1pks-nrps	8	40	5	NG-391*, Fengycin (*fenA–E*), Surfacin (*surAA-AC*)	ABCC	**
18	t1pks-nrps	14	27	14	Gramicidin (*grsA-B*), Fengycin (*fenA–E*), Surfacin (*surAA-AC*)	-	*
19	t1pks-nrps	20	13	3	Lovastatin (*LOVB, LOVF*), Fengycin (*fenA-E*), Surfactin (*srfAA-AC*)	ABCB	*
20	terpene	12	24	11	Fernesyl-diphosphate (*FDFT1*)	-	**
21	terpene	17	12	13	Lupeol (*LUP1,2,4,5*), Arabidiol (*PEN1*), Tinucalladienol (*PEN3*), seco-amyrin (*PEN6*)	-	*

ID	**Type of GCs**	**Pi**	**Ru**	**Pu**	**Secondary metabolic pathways detected in DNA sequence homology**	**ATP-binding cassettes**	**Identity level**
22	terpene	-	-	15	-	-	-
23	others	15	42	18	Fengycin (*fenA, C, E*), Surfacin (srfAA-AB)	ABCF	**
24	others	6	32	6	Kinesin (KIDFC1-3)	-	*
25	others	10	5	16	Abscisic aldehyde (AAO1-4), Fengycin (*fenA, B, E*), Surfactin (*srfAA, AB*)	ABCB	*
26	others	16	33	4	Histidinol (*hisD, IE*)	-	*
27	others	21	28	1	-	-	*
28	cf-putative	-	1	-	-	-	-
29	cf-putative	-	3	-	Palmitin (*ZDHHC*)	-	-
30	cf-putative	-	6	-	Palmitin (*ZDHHC4)*	-	-
31	cf-putative	-	7	-	-	-	-
32	cf-putative	-	8	-	-	-	-
33	cf-putative	-	9	-	-	-	-
34	cf-putative	-	11	-	-	-	-
35	cf-putative	-	14	-	Serine, Threonine (*PP1C, 2C, 4C, 6C*)	-	-
36	cf-putative	-	15	-	-	ABCG2	-
37	cf-putative	-	16	-	-	-	-
38	cf-putative	-	17	-	-	-	-
39	cf-putative	-	18	-	Mannan (*ANPI, MNN9*)	-	-
40	cf-putative	-	20	-	-	-	-
41	cf-putative	-	25	-	-	-	-
42	cf-putative	-	29	-	-	-	-
43	cf-putative	-	31	-	-	-	-
44	cf-putative	-	34	-	-	-	-
45	cf-putative	-	36	-	-	-	-
46	cf-putative	-	37	-	-	-	-
47	cf-putative	-	38	-	-	-	-

ID	Type of GCs	Pi	Ru	Pu	Secondary metabolic pathways detected in DNA sequence homology	ATP-binding cassettes	Identity level
48	cf-putative	-	39	-	Lovastatin (*LOVB-G*), Phthiocerol/Phenolphthiocerol (*ppsA,C*)	-	-
49	cf-putative	-	44	-	-	-	-
50	cf-putative	-	45	-	-	-	-
51	cf-putative	-	46	-	-	-	-
52	cf-putative	-	47	-	-	-	-
53	cf_fatty_acid	-	26	-	-	-	-
54	cf_fatty_acid	-	35	-	Fatty acid (*FAS1,2*)	-	-

Table 2 Comparison of six *M. purpureus*, *M. ruber*, and *M. pilosus* strains based on *Monascus* azaphilone pigment polyketide synthase gene. Mutated nucleotides of the target gene were extracted and summarized in the table.

	30	57	88	108	117	123	161	216	219	288	291	321	369	376	392	394	423	442	447	466	471	478	483	516	578	589	606	627	648	663	685	703	748	774	802	820	834	860	939	966	1022	1074	1085	1112	1152	1172	1180	1186	1188	1193
M .pilosus NBRC 4520	A	A	A	T	T	G	G	G	A	T	T	A	G	C	G	A	T	A	A	A	A	A	T	C	C	T	A	A	C	T	A	A	T	T	T	C	C	T	T	A	A	T	C	A	A	T	A	C	T	T
M .pilosus NCBI	A	A	A	T	T	G	G	G	A	T	T	A	G	C	G	A	T	A	A	A	A	A	T	C	C	T	A	A	C	T	A	A	T	T	T	C	C	T	T	A	A	T	C	A	A	T	A	C	T	T
M. ruber NRRP 1597	A	A	A	T	T	G	G	G	A	T	T	A	G	C	G	A	T	A	A	A	A	A	T	C	C	T	A	A	C	T	A	A	T	T	T	C	C	T	T	A	A	T	C	A	A	T	A	C	T	T
M. ruber NBRC 4483	A	A	A	T	T	G	G	G	A	T	T	A	G	C	G	A	T	A	A	A	A	A	T	C	C	T	A	A	C	T	A	A	T	T	T	C	C	T	T	A	A	T	C	A	A	T	A	C	T	T
M. ruber JF83291.6	C	G	G	C	C	C	A	A	G	C	C	G	A	G	A	G	C	G	G	C	G	G	C	T	T	G	G	G	T	C	G	G	C	G	C	C	A	C	C	G	G	C	T	G	G	C	G	T	C	C
M. purpureus NRRP 1596	C	G	G	C	C	C	A	A	G	C	C	G	A	G	A	G	C	G	G	C	G	G	C	T	T	G	G	G	T	C	G	G	C	G	C	A	A	C	C	G	G	C	T	G	G	C	G	T	C	C
M. purpureus NBRC 4478	C	G	G	C	C	C	A	A	G	C	C	G	A	G	A	G	C	G	G	C	G	G	C	T	T	G	G	G	T	C	G	G	C	G	C	A	A	C	C	G	G	C	T	G	G	C	G	T	C	C

	1204	1209	1235	1236	1241	1244	1274	1297	1305	1307	1318	1349	1352	1371	1376	1385	1391	1394	1395	1432	1487	1508	1517	1541	1601	1631	1658	1742	1793	1827	1865	1937	1973	1981	2030	2141	2187	2202	2219	2267	2273	2291	2375	2378	2405	2465	2483	2510	2532	2655
M .pilosus NBRC 4520	T	A	G	A	T	T	T	A	A	G	A	A	G	G	T	T	A	T	A	T	C	G	C	C	T	T	A	T	T	C	T	A	A	A	T	A	G	T	G	T	A	A	G	C	T	T	T	G	C	T
M .pilosus NCBI	T	A	G	A	T	T	T	A	A	G	A	A	G	G	T	T	A	T	A	T	C	G	C	C	T	T	A	T	T	C	T	A	A	A	T	A	G	T	G	T	A	A	G	C	T	T	T	G	C	T
M. ruber NRRP 1597	T	A	G	A	T	T	T	A	A	G	A	A	G	G	T	T	A	T	A	T	C	G	C	C	T	T	A	T	T	C	T	A	A	A	T	A	G	T	G	T	A	A	G	C	T	T	T	G	C	T
M. ruber NBRC 4483	T	A	G	A	T	T	T	A	A	G	A	A	G	G	T	T	A	T	A	T	C	G	C	C	T	T	A	T	T	C	T	A	A	A	T	A	G	T	G	T	A	A	G	C	T	T	T	G	C	T
M. ruber JF83291.6	C	G	C	G	C	C	G	G	G	C	G	G	A	A	C	C	G	A	C	C	T	A	T	T	C	G	G	C	C	A	C	T	G	G	C	C	A	C	A	G	G	G	A	T	G	C	C	C	T	G
M. purpureus NRRP 1596	C	G	C	G	C	C	G	G	G	C	G	G	A	A	C	C	G	A	C	C	T	A	T	T	C	G	G	C	C	A	C	T	G	G	C	C	A	C	A	G	G	G	A	T	G	C	C	C	T	G
M. purpureus NBRC 4478	C	G	C	G	C	C	G	G	G	C	G	G	A	A	C	C	G	A	C	C	T	A	T	T	C	G	G	C	C	A	C	T	G	G	C	C	A	C	A	G	G	G	A	T	G	C	C	C	T	G

	2681	2699	2708	2714	2720	2723	2747	2819	2867	2888	2929	2930	2933	2960	2970	2999	3086	3089	3178	3201	3264	3265	3266	3332	3360	3368	3381	3386	3457	3458	3483	3486	3531	3551	3602	3624	3626	3633	3650	3657	3671	3677	3688	3722	3764	3798	3800	3848	3907	3918
M .pilosus NBRC 4520	C	T	C	T	T	A	A	T	T	T	T	T	T	G	A	A	A	G	A	A	-	-	-	A	A	T	T	T	A	A	A	C	G	G	A	G	T	C	A	G	T	T	T	T	C	T	C	G	T	G
M .pilosus NCBI	C	T	C	T	T	A	A	T	T	T	T	T	T	G	A	A	A	G	A	A	-	-	-	A	A	T	T	T	A	A	A	C	G	G	A	G	T	C	A	G	T	T	T	T	C	T	C	G	T	G
M. ruber NRRP 1597	C	T	C	T	T	A	A	T	T	T	T	T	T	G	A	A	A	G	A	A	-	-	-	A	A	T	T	T	A	A	A	C	G	G	A	G	T	C	A	G	T	T	T	T	C	T	C	G	T	G
M. ruber NBRC 4483	C	T	C	T	T	A	A	T	T	T	T	T	T	G	A	A	A	G	A	A	-	-	-	A	A	T	T	T	A	A	A	C	G	G	A	G	T	C	A	G	T	T	T	T	C	T	C	G	T	G
M. ruber JF83291.6	T	C	T	C	C	G	G	C	C	C	C	C	C	T	T	G	C	A	C	G	G	A	C	G	G	C	G	C	G	G	G	T	A	A	G	A	C	T	C	A	C	G	C	C	T	C	T	A	C	A
M. purpureus NRRP 1596	T	C	T	C	C	G	G	C	C	C	C	C	C	T	T	G	C	A	C	G	G	A	C	G	G	C	G	C	G	G	G	T	A	A	G	A	C	T	C	A	C	G	C	C	T	C	T	A	C	A
M. purpureus NBRC 4478	T	C	T	C	C	G	G	C	C	C	C	C	C	T	T	G	C	A	C	G	G	A	C	G	G	C	G	C	G	G	G	T	A	A	G	A	C	T	C	A	C	G	C	C	T	C	T	A	C	A

	3919	3936	3953	4025	4058	4088	4151	4169	4248	4310	4373	4425	4481	4484	4592	4660	4670	4673	4775	4785	4808	4912	4927	4937	4958	4974	4985	4987	5021	5035	5036	5182	5194	5198	5230	5234	5276	5281	5291	5292	5293	5294	5301	5302	5363	5417	5442	5448	5471	5489
M .pilosus NBRC 4520	T	A	A	C	T	C	A	A	T	G	T	A	A	T	A	T	C	T	A	A	T	T	G	T	A	A	T	C	C	C	C	T	G	A	G	T	T	T	T	-	-	-	G	G	C	C	T	A	A	A
M .pilosus NCBI	T	A	A	C	T	C	A	A	T	G	T	A	A	T	A	T	C	T	A	A	T	T	G	T	A	A	T	C	C	C	C	T	G	A	G	T	T	T	T	-	-	-	G	G	C	C	T	A	A	A
M. ruber NRRP 1597	T	A	A	C	T	C	A	A	T	G	T	A	A	T	A	T	C	T	A	A	T	T	G	T	A	A	T	C	C	C	C	T	G	A	G	T	T	T	T	-	-	-	G	G	C	C	T	A	A	A
M. ruber NBRC 4483	T	A	A	C	T	C	A	A	T	G	T	A	A	T	A	T	C	T	A	A	T	T	G	T	A	A	T	C	C	C	C	T	A	A	G	T	T	T	T	-	-	-	G	G	C	C	T	A	A	A
M. ruber JF83291.6	C	G	G	T	C	G	G	T	C	C	C	T	G	C	T	C	T	C	C	C	C	C	A	G	G	G	C	T	T	T	T	G	G	G	A	C	C	G	A	G	A	G	A	A	T	T	G	G	G	G
M. purpureus NRRP 1596	C	G	G	T	C	G	G	T	C	C	C	T	G	C	T	C	T	C	C	C	C	C	A	G	G	G	C	T	T	T	T	G	G	G	A	C	C	G	A	G	A	G	A	A	T	T	G	G	G	G
M. purpureus NBRC 4478	C	G	G	T	C	G	G	T	C	C	C	T	G	C	T	C	T	C	C	C	C	C	A	G	G	G	C	T	T	T	T	G	G	G	A	C	C	G	A	G	A	G	A	A	T	T	G	G	G	G

	5492	5495	5500	5501	5527	5552	5568	5599	5615	5618	5634	5674	5714	5726	5780	5795	5807	5819	5882	5906	5987	6021	6028	6041	6149	6167	6173	6200	5021	6236	6248	6329	6347	6383	6389	6521	6569	6572	6596	6641	6759	6914	6928	7002	7006	7007	7008	7019	7024	7082
M .pilosus NBRC 4520	C	A	T	A	C	T	A	T	C	C	G	G	C	A	T	C	C	A	G	T	A	T	T	C	T	A	T	A	C	T	T	T	A	A	T	C	C	G	T	G	T	T	T	A	C	A	T	T	G	C
M .pilosus NCBI	C	A	T	A	C	T	A	T	C	C	G	G	C	A	T	C	C	A	G	T	A	T	T	C	T	A	T	A	C	T	T	T	A	A	T	C	C	G	T	G	T	T	T	A	C	A	T	T	G	C
M. ruber NRRP 1597	C	A	T	A	C	T	A	T	C	C	G	G	C	A	T	C	C	A	G	T	A	T	T	C	T	A	T	A	C	T	T	T	A	A	T	C	C	G	T	G	T	T	T	A	C	A	T	T	G	C
M. ruber NBRC 4483	C	A	T	A	C	T	A	T	C	C	G	G	C	A	T	C	C	A	G	T	A	T	T	C	T	A	T	A	C	T	T	T	A	A	T	C	C	G	T	G	T	T	T	A	C	A	T	T	G	C
M. ruber JF83291.6	T	G	G	G	T	C	G	C	G	T	A	C	G	G	A	T	T	G	A	G	G	C	C	T	A	G	C	G	T	C	C	C	C	G	C	T	T	C	C	C	C	C	C	G	-	-	-	C	A	T
M. purpureus NRRP 1596	T	G	G	G	T	C	G	C	G	T	A	C	G	G	A	T	T	G	A	G	G	C	C	T	A	G	C	G	T	C	C	C	C	G	C	T	T	C	C	C	C	C	C	G	-	-	-	C	A	T
M. purpureus NBRC 4478	T	G	G	G	T	C	G	C	G	T	A	C	G	G	A	T	T	G	A	G	G	C	C	T	A	G	C	G	T	C	C	C	C	G	C	T	T	C	C	C	C	C	C	G	-	-	-	C	A	T

	7189	7191	7241	7270	7316	7325	7343	7397	7415	7421	7445	7457	7556	7583	7631	7676	7692	7769	7853	7874	7976	8043	8044	8083	8105	8107
M .pilosus NBRC 4520	A	A	G	A	C	T	C	A	T	C	T	A	T	A	G	G	A	C	C	T	T	G	T	T	A	A
M .pilosus NCBI	A	A	G	A	C	T	C	A	T	C	T	A	T	A	G	G	A	C	C	T	T	G	T	T	A	A
M. ruber NRRP 1597	A	A	G	A	C	T	C	A	T	C	T	A	T	A	G	G	A	C	C	T	T	G	T	T	A	A
M. ruber NBRC 4483	A	A	G	A	C	T	C	A	T	C	T	A	T	A	G	G	A	C	C	T	T	G	T	T	A	A
M. ruber JF83291.6	G	C	C	G	T	C	T	C	C	T	C	G	C	G	A	A	G	T	G	C	G	C	C	G	G	G
M. purpureus NRRP 1596	G	C	C	G	T	C	T	C	C	T	C	G	C	G	A	A	G	T	G	C	G	C	C	G	G	G
M. purpureus NBRC 4478	G	C	C	G	T	C	T	C	C	T	C	G	C	G	A	A	G	T	G	C	G	C	C	G	G	G

Table 3 Homologous peptide sequences of the citrinin biosynthetic genes in three *Monascus* species.

Protein	*M. pilosus* NBRC 4520		*M. ruber* NBRC 4483		*M. purpureus* NBRC 4478		Reference peptide sequence
	length	(%)	length	(%)	length	(%)	length
CitA (*mpl1*)	249	79.6	249	79.6	312	99.7	313
CitB (*mpl2*)	229	69.6	229	69.6	328	99.7	329
CitD (*mpl4*)	458	91.4	458	91.4	500	99.8	501
CitE (*mpl6*)	231	79.1	231	79.1	284	97.3	292
CitC (*mpl7*)	532	85.5	532	85.5	618	99.4	622
CitS (*pksCT*)	1525	58.8	1525	58.8	2396	92.4	2593

Table 4 Homologous peptide sequences of the monacolin biosynthetic genes in three *Monascus* species.

	M. pilosus NBRC 4520		*M. ruber* NBRC 4483		*M. purpureus* NBRC 4478		Reference peptide sequence
Protein	length	(%)	length	(%)	length	(%)	length
MokA	1910	100	1910	100	1772	92.8	1910
MokB	1947	100	1947	100	1501	77.1	1947
MokC	339	100	339	100	28	8.3	339
MokD	263	100	263	100	154	58.6	263
MokE	265	100	265	100	231	87.2	265
MokF	258	100	258	100	167	64.7	258
MokG	996	99.0	977	100	322	32.3	977
MokH	487	100	487	100	42	8.6	487
MokI	174	100	174	100	107	61.5	174