

Table S1 – Micro-synteny analysis between genes in the 20 longest PST-130 contigs and their putative orthologs in PGTG

PST-130 contig	Length	PGTG contig AACC0100	PST coordinates	% PST Contig	PGTG proteins (PGTG_) ^a	No. common genes ^b	Super contig No. ^c	Comments
8308	49498	1263	4656-44745	86%	06750 , 52, 49, 48, 47, 46, 45, 44, 43	9	15	1 multigene family
6952	39939	3515	4913-34166	73%	17475 , 82, 81, 80, 79, 78	6	10	
		3750	4913-34166	73%	18527 , 31, 32, 33, 34, 35	6	17	Duplicated PGTG region
7251	39628	1311 1326	870-10776 25550-33655	45%	06818 , 20, 22 / 55, 58	5	17	
7806	37324	1487 1482 1486	11-848 14684-17049 17707-22646	22%	07698 , 97, 95, 93 / 83 / 82, 81, 80	8	20	
7242	37091	0045 0046 0047	2532-2566 8610-29864 31324-34905	67%	00398 / 00401 , 02, 06, 07, 08, 09 / 16	8	1	
7402	36035	0221 0220 0219	11094-23115 28378-29074 33433-33520	34%	00879 , 78, 76 / 72 / 68	5	3	
8617	35916	3253	3840-32211	79%	16684 , 83, 80, 79, 78, 77, 76, 75, 74	9	74	
20459	35740	0414	486-30724	85%	02069 , 73, 74, 77	4	5	
7057	34936	0617 0620 0623	3799-4591 19969-22905 28260-34382	50%	03701 / 08, 09 / 27, 28, 29	6	7	
13600	34906	2406	2562-13395	31%	13265 , 64, 63, 62	4	43	
		0390	19187-19576		02861		4	Not syntenic ^d
		1580	20773-21157		08875		22	Not syntenic
		3724 3768 3725	23514-26577 27273-29880 31944-33445	21%	18238 , 42 / 45 / 48	4	103	
7101	34773	0559 1931	3731-30508 33328-34461	80%	03381 , 82, 83, 84, 86, 87 / 88, 89	8	5	
7598	34231	2324/5	2399-29991	81%	12949 , 50, 51, 56, 39, 57, 35, 59, 60	9	40	
8613	33720	0033 0034	27688-32436 5089-22558	66%	00295 / 00313 , 16, 15	4	1	
7582	33253	0329 0331	2010-12997 28970-31548	41%	02520 , 19, 18 / 16	4	4	
17625	32924	0310 1583	11665-12077 14171-30709	52%	08915 / 13, 07	3	22	

7330	32453	3539 3544 3550	2514-2794 2568-7365 17699-28388	49%	18037 / 48, 49, 54 / 59, 60, 61, 62, 64	9	92	
7170	32391	0886 0885	3425-25416 29853-29215	66%	04952 , 50, 48, 47, 46 / 42	6	10	
		0923	31023-32017	3%	05469 , 68	2	11	
6894	31719	1834 1833	751-1060 4760-25274	66%	10580 / 69, 67, 66,64, 62, 60	7	28	
7946	31602	3551	2010-2684		18069		92	Not syntenic
		2278	6272-8077					
		2279 2280 2282	14342-14497 17354-23419 29588-31601	32%	12265 / 70 / 71 / 77	4	38	
7043	31566	0344	1094-28034	85%	02621 , 14, 13, 11, 05, 04	6	4	

^a Adjacent protein numbers usually correspond to adjacent annotated genes. Numbers separated by comas start with the same numbers in bold. Most of the large PST contigs have large stretches that are syntenic to single PGTG contigs. In several cases when there are multiple PGTG contigs they correspond to the same PGTG supercontig. Most of the corresponding PGTG proteins that match a PST contig also have consecutive or close numbers, also indicating synteny. Altered orders indicate rearrangements.

^b The number of syntenic genes includes those genes that have undergone rearrangements within the syntenic group of genes.

^c Supercontigs are from PGTG genome assembly (01/11/2011, www.broadinstitute.org).

^d Single gene matches in a different supercontig were considered not syntenic.