

Table S1. Quality-control processing of ESTs.

Numbers labeled in bold italics indicate dbEST from cultivars that did not meet the criteria listed in the Materials and Methods.

plant	strain	dbEST <sup>a</sup>	Used dbEST <sup>b</sup>	EST size <sup>c</sup>	Cluster# <sup>d</sup>	%Flowers or berries <sup>e</sup>
<i>Arabidopsis thaliana</i>	columbia	215633	111142	440	2751	7.8
<i>Arabidopsis thaliana</i>	wassilewskija	49667	49667	470	2522	0.0
<i>Glycine max</i>	williams	94752	91691	444	2504	0.0
<i>Glycine max</i>	williams 82	82674	80218	492	2873	17.4
<i>Gossypium raimondii</i>	no strain	63117	63117	<b>714</b>	2924	0.0
<i>Hordeum vulgare</i>	barke	114423	96462	515	2571	3.9
<i>Hordeum vulgare</i>	haruna nijo	86597	86597	472	2905	0.0
<i>Hordeum vulgare</i>	morex	52428	42588	<b>711</b>	2294	0.0
<i>Ipomoea nil</i>	tokyo-kokei	61869	61869	601	<b>3926</b>	<b>41.7</b>
<i>Lactuca sativa</i>	salinas	80144	80144	657	3540	0.0
<i>Lotus japonicus</i>	gifu b-129	70082	48654	446	2388	3.5
<i>Lotus japonicus</i>	miyakojima mg-20	75268	75268	410	3023	<b>20.7</b>
<i>Malus x domestica</i>	goldrush	91843	76574	457	3160	<b>28.9</b>
<i>Malus x domestica</i>	royal gala	117790	117790	467	2799	5.7
<i>Medicago truncatula</i>	a17	71611	66752	586	3254	6.6
<i>Oryza sativa (indica)</i>	pa64s	50882	50882	470	2813	<b>86.3</b>
<i>Oryza sativa (japonica)</i>	nackdong	61830	<b>28527</b>	460		18.6
<i>Oryza sativa (japonica)</i>	nipponbare	893777	893777	423	3183	14.2
<i>Solanum lycopersicum</i>	micro-tom	92074	92074	541	3737	0.0
<i>Solanum lycopersicum</i>	ta496	108421	89186	522	3179	16.7
<i>Sorghum bicolor</i>	btx623	160752	128043	535	2810	0.0
	cabernet					
<i>Vitis vinifera</i>	sauvignon	158890	134202	582	2920	<b>58.4</b>
<i>Vitis vinifera</i>	chardonnay	56596	<b>35660</b>	541	2985	5.1
<i>Zea mays</i>	b73	645237	645237	518	2102	0.0
<i>Zea mays</i>	f2	87344	58150	409	2714	0.0
<i>Zea mays</i>	oh43	54168	54168	457	2836	0.0

<sup>a</sup>EST were download from NBCI 23/12/06 contained within 711 libraries. dbESTs are ESTs from the specific cultivar with at least 40,000 ESTs, that are more than 99bp and do not include; poly A, vector contamination or more than 10% Ns (Materials and Methods; step i, ii and iii).

<sup>b</sup>ESTs from dbEST column not including libraries that contain more than 1% ESTs that include more than one indel. Cultivars that do not include at least 40,000 ESTs after processing the libraries, are marked in bold italics and were not used (Materials and Methods; step iv).

<sup>c</sup> Average ESTs size in each plant group. Average ESTs size, marked in bold italics are more than 2 standard deviation from the average ESTs sizes (Materials and Methods; step v).

<sup>d</sup>Used ESTs were grouped into clusters. Cluster numbers, that are more than 2 standard deviation from the average cluster number, are marked in bold italic (Materials and Methods; step vi).

<sup>e</sup>ESTs from libraries that contain only flowers or berries were counted. Cultivars in which those ESTs are more than 20% from all the ESTs are marked in bold italics (Materials and Methods; step vii).