PLOS ONE

Whole-Genome Quantitative Trait Locus Mapping Reveals Major Role of Epistasis on Yield of Rice

Anhui Huang[1], Shizhong Xu[2], Xiaodong Cai[1]*

1 Department of Electrical and Computer Engineering, University of Miami, Coral Gables, Florida, United State of America, **2** Department of Botany and Plant Sciences, University of California Riverside, Riverside, California, United State of America

Abstract

Although rice yield has been doubled in most parts of the world since 1960s, thanks to the advancements in breeding technologies, the biological mechanisms controlling yield are largely unknown. To understand the genetic basis of rice yield, a number of quantitative trait locus (QTL) mapping studies have been carried out, but whole-genome QTL mapping incorporating all interaction effects is still lacking. In this paper, we exploited whole-genome markers of an immortalized F_2 population derived from an elite rice hybrid to perform QTL mapping for rice yield characterized by yield per plant and three yield component traits. Our QTL model includes additive and dominance main effects of 1,619 markers and all pair-wise interactions, with a total of more than 5 million possible effects. The QTL mapping identified 54, 5, 28 and 4 significant effects involving 103, 9, 52 and 7 QTLs for the four traits, namely the number of panicles per plant, the number of grains per panicle, grain weight, and yield per plant. Most identified QTLs are involved in digenic interactions. An extensive literature survey of experimentally characterized genes related to crop yield shows that 19 of 54 effects, 4 of 5 effects, 12 of 28 effects and 2 of 4 effects for the four traits, respectively, involve at least one QTL that locates within 2 cM distance to at least one yield-related gene. This study not only reveals the major role of epistasis influencing rice yield, but also provides a set of candidate genetic loci for further experimental investigation.

Introduction

Given the paramount importance in sustaining food demanding, great efforts have been made in large scale genetic research and extensive breeding programs in almost all rice (*Oryza sativa* L.) producing countries [1,2]. Gains in rice yield in recent decades are mainly owed to advancements in breeding technologies including selection of cultivars with higher productivity and significant increase of agricultural inputs such as fertilizers and insecticides [3]. While global environmental degradation has limited further yield increase through more agricultural inputs, studying the underlying biological processes of rice yield, and transferring the knowledge gains into improvement in breeding and agronomic productivity have become the key for further increase of food production [4].

Rice yield is determined by several factors including the number of panicles per plant, the number of grains per panicle and grain weight. These component traits and the overall yield per plant exhibit continuous variation since they are influenced by multiple genetic factors named quantitative trait loci (QTLs) and other environmental factors. Genetic markers such as restriction fragment length polymorphisms (RFLPs) [5] and simple sequence repeats (SSRs) [6] have been utilized to identify QTLs for understanding genetic basis controlling rice yield [7–11]. A recent study on QTL mapping for rice yield derived a high density single nucleotide polymorphism (SNP) bin map from genomic sequences

obtained using deep sequencing technology, and demonstrated that such high density SNP bin map enabled to identify more QTLs with higher location precision than the traditional approach based on RFLP and SSR markers [12]. However, these studies attempted to identify QTLs individually via single interval mapping [5] or composite interval mapping with a small scan window [13], which had limited power of detection, given that many agronomic traits are controlled simultaneously by multiple QTLs and influenced by environmental factors [14,15].

Whole-genome marker QTL mapping employs a multiple QTL model that includes all available markers and evaluates effects of these markers simultaneously [16–18]. Such approach overcomes the limitations of the traditional single marker-based QTL mapping methods [16]. However, when genetic interactions are considered, a multiple QTL model can have a huge number of variables, which makes model inference very challenging. Early methods for multiple QTL mapping usually rely on Markov chain Monte Carlo (MCMC) simulation to fit a Bayesian model [16–20], which is computationally intensive and unpractical when a large number of markers are considered. Recently, more efficient and accurate methods have been developed [21,22], which make whole-genome marker QTL mapping feasible. With whole-genome marker QTL mapping considering main effects and interactions of all additive and dominance effects simultaneously, contributions of numerous genetic effects to rice yield can be assessed.

Table 1. Estimated QTL effects from the full model for the number of panicles per plant.

Loci(i, j)[a]	$\hat{\beta}(s_{\hat{\beta}})$[b]	p-value[c]	$\hat{h}_j^2$[d]	Loci(i, j)	$\hat{\beta}(s_{\hat{\beta}})$	p-value	$\hat{h}_j^2$
(757_add, 757_add)	−0.12(0.04)	2.16×10^{-3}	0.0030	(104_dom, 732_add)	−0.11(0.05)	9.54×10^{-3}	0.0012
7_add, 220_dom)	0.20(0.06)	2.39×10^{-4}	0.0032	(186_dom, 735_add)	−0.20(0.06)	1.72×10^{-4}	0.0039
(10_add, 887_dom)	−0.25(0.05)	3.78×10^{-6}	0.0061	(518_dom, 759_add)	−0.27(0.06)	1.64×10^{-6}	0.0075
(18_add, 1407_dom)	0.28(0.06)	1.45×10^{-6}	0.0080	(220_dom, 784_add)	0.23(0.06)	2.38×10^{-5}	0.0045
(20_add, 1026_dom)	0.27(0.06)	1.82×10^{-6}	0.0060	(561_dom, 828_add)	−0.36(0.05)	4.88×10^{-12}	0.0140
(44_add, 532_dom)	−0.27(0.05)	1.16×10^{-7}	0.0071	(861_add, 918_add)	0.41(0.05)	1.11×10^{-15}	0.0182
(69_add, 913_dom)	0.21(0.05)	2.39×10^{-5}	0.0040	(904_add, 1113_dom)	0.33(0.05)	3.50×10^{-10}	0.0098
(123_add, 1132_add)	0.23(0.05)	7.78×10^{-7}	0.0053	(213_dom, 929_add)	0.27(0.05)	4.85×10^{-8}	0.0076
(166_add, (684_add)	0.46(0.04)	$<10^{-15}$	0.0279	(967_add, 1515_add)	0.20(0.05)	2.91×10^{-5}	0.0040
(186_add, 1372_add)	−0.11(0.04)	2.86×10^{-3}	0.0013	(908_dom, 994_add)	0.90(0.05)	$<10^{-15}$	0.0782
(192_add, 580_add)	−0.11(0.04)	8.35×10^{-3}	0.0013	(1026_add, 1173_add)	0.21(0.05)	9.05×10^{-6}	0.0044
(199_add, 782_dom)	−0.08(0.03)	2.53×10^{-3}	0.0006	(1037_add, 1510_add)	0.14(0.05)	1.13×10^{-3}	0.0018
(208_add, 309_add)	−0.31(0.05)	1.50×10^{-9}	0.0092	(1089_dom, 1096_add)	0.34(0.12)	2.17×10^{-3}	0.0016
(227_add, 364_dom)	−0.36(0.05)	3.18×10^{-12}	0.0145	(1119_add, 1471_add)	−0.19(0.04)	1.29×10^{-5}	0.0045
(244_add, 1303_dom)	−0.11(0.04)	5.10×10^{-3}	0.0012	(229_dom, 1160_add)	−0.11(0.04)	5.70×10^{-3}	0.0012
(249_add, 417_dom)	0.12(0.04)	3.14×10^{-3}	0.0015	(1208_add, 1583_dom)	−0.14(0.05)	3.20×10^{-3}	0.0015
(333_add, 991_add)	0.24(0.05)	3.92×10^{-7}	0.0057	(64_dom, 1223_add)	−0.43(0.05)	2.22×10^{-15}	0.0191
(335_add, 372_add)	0.20(0.05)	2.05×10^{-5}	0.0041	(1237_add, 1370_add)	0.54(0.05)	$<10^{-15}$	0.0279
(349_add, 1425_dom)	−0.23(0.05)	3.20×10^{-6}	0.0060	(1334_add, 1576_add)	0.22(0.05)	2.94×10^{-6}	0.0049
(354_add, 358_dom)	−0.50(0.05)	$<10^{-15}$	0.0233	(408_dom, 1356_add)	−0.73(0.05)	$<10^{-15}$	0.0488
(371_dom, 381_add)	0.37(0.05)	6.01×10^{-11}	0.0113	(1065_dom, 1394_add)	−0.17(0.05)	1.55×10^{-4}	0.0026
(421_add, 1079_add)	−0.15(0.04)	1.04×10^{-4}	0.0023	(981_dom, 1558_add)	−0.79(0.05)	$<10^{-15}$	0.0735
(456_add, 1282_add)	0.38(0.05)	9.99×10^{-15}	0.0167	(1094_dom, 1558_add)	0.21(0.05)	4.05×10^{-5}	0.0046
(517_add, 1346_add)	−0.10(0.04)	8.46×10^{-3}	0.0009	(1217_dom,1615_add)	0.37(0.05)	5.88×10^{-13}	0.0144
(520_add,595_dom)	−0.37(0.05)	4.03×10^{-11}	0.0109	(54_dom, 1117_dom)	0.27(0.05)	3.96×10^{-7}	0.0052
(15_dom, 534_add)	0.47(0.05)	$<10^{-15}$	0.0246	(627_dom, 681_dom)	−0.25(0.05)	1.05×10^{-6}	0.0048
(649_add, 1364_add)	−0.20(0.04)	3.80×10^{-6}	0.0045	(786_dom, 810_dom)	0.15(0.05)	1.11×10^{-3}	0.0021
Parameter(s)				$a=0.5$, $b=0.5$			
μ				0.0035			
σ_0^2				0.1444			
$\hat{h}^2$				0.9405			

[a]add: additive effect; dom: dominance effect. If i equals j, then it is a main effect, otherwise, it is an interaction between locus i and locus j. Total number of effects is 112, only 54 effects with a p-value ≤ 0.01 are listed in this table.
[b]The estimated marker effect is denoted by $\hat{\beta}$ and the standard deviation is denoted by $s_{\hat{\beta}}$.
[c]P-value is obtained via t-test.
[d]Phenotypic variation explained.
doi:10.1371/journal.pone.0087330.t001

In this study, we applied our empirical Bayesian least absolute shrinkage and selection operator (EBlasso) method [21,22] to whole-genome QTL mapping for an elite *indica* rice hybrid, Shanyou 63 [7,23]. Our EBlasso model includes additive and dominance main effects of 1,619 markers, all additive × additive interactions, additive × dominance interactions, dominance × additive interactions, and dominance × dominance interactions, with a total of more than 5 million possible effects. The quantitative traits considered in this study include yield per plant and three yield component traits, namely the number of panicles per plant, the number of grains per panicle and grain weight. We will demonstrate that our EBlasso identifies a number of QTLs, most of which are involved in digenic interactions, and coincide with or are close to experimentally investigated genes related to yield.

Results

Four quantitative traits including three rice yield component traits (the number of panicles per plant, the number of grains per panicle and grain weight) and overall yield per plant were analyzed using the EBlasso method. The full QTL model includes main additive and dominance effects of 1,619 markers and all their pair-wise interactions, with a total of $k=5,242,322$ variables (see the Materials and Methods section for the genetic map). To understand the performance gain of the full model, we also performed QTL mapping for the four traits with a QTL model including $k=3,238$ main effects, which is referred to as the main effect model.

We estimated the phenotypic variance explained by a particular QTL j as $h_j^2 = \dfrac{\hat{\beta}_j^2 \mathrm{var}(\boldsymbol{x_j})}{\sigma_y^2}$, $j=1, 2, \ldots, k'$, where $\mathrm{var}(\boldsymbol{x_j})$ is the
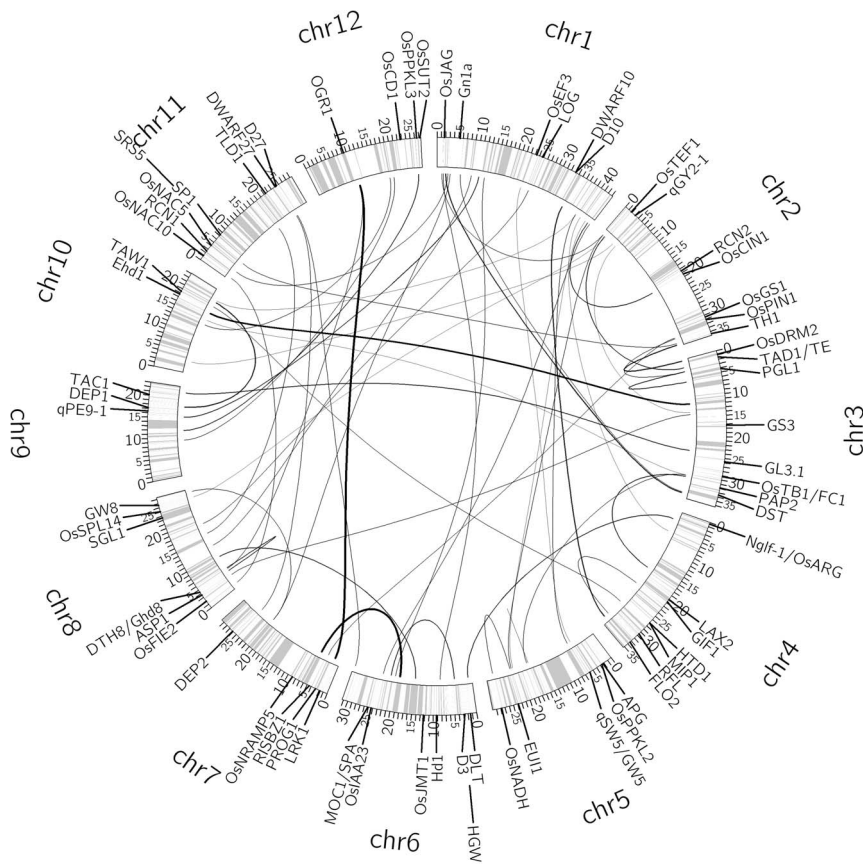
Figure 1. Interaction network of 103 QTLs for the number of panicles per plant. The circle shows the bin map and columns indicate position of the makers (ticks in million base pairs). The thickness of a link is proportional to the strength of the interaction effect. A short straight line indicates a main effect. Molecularly characterized genes related to yield are also labeled in the appropriate positions of the genome.
doi:10.1371/journal.pone.0087330.g001

variance of the coefficient of QTL j and the total phenotypic variance σ_y^2 was estimated from the data. To estimate the total variance explained by all identified QTLs, we refitted the data to an ordinary linear regression model that includes variables corresponding to the identified QTLs. The phenotypic values were predicted from the linear regression model as $\hat{y}$, and the total phenotypic variance explained by all identified QTLs was calculated as

$$h^2 = \frac{\mathrm{var}(\hat{y})}{\mathrm{var}(y)} = \frac{\mathrm{var}(\hat{y})}{\sigma_y^2} \qquad (1)$$

.

QTL mapping for the number of panicles per plant

The three-step cross validation (CV) procedure (detailed in the Materials and Methods section) for the full model identified the optimal pair of parameters as $(a, b) = (0.5, 0.5)$ (Table S1 in File S1). Using the optimal values of (a, b), the EBlasso algorithm shrunk most of k variables to zero and yielded a QTL model with 111 nonzero effects. The statistical test, described in the Materials and Methods section, for each nonzero effect identified 54 significant effects at a p-value ≤ 0.01 (Table 1). Among them, one was main additive effect, 18 were additive × additive interaction, 32 were additive ×dominance interaction, and three were dominance × dominance interaction. The 54 effects involved 103 QTLs and explained 94.05% of the total phenotypic variance.

Table 2. Estimated QTL effects from the main effect model for the number of panicles per plant.

locus[a]	$\hat{\beta}(s_{\hat{\beta}})$[b]	p-value[c]	$\hat{h}_j^2$[d]
3_add	0.22(0.09)	6.86×10^{-3}	0.0088
228_add	−0.24(0.09)	2.98×10^{-3}	0.0125
353_add	−0.24(0.09)	2.68×10^{-3}	0.0126
757_add	−0.54(0.10)	4.18×10^{-8}	0.0625
818_add	0.40(0.10)	4.82×10^{-5}	0.0300
908_add	0.31(0.09)	4.78×10^{-4}	0.0206
994_add	0.54(0.11)	3.53×10^{-7}	0.0524
1363_add	−0.26(0.09)	2.29×10^{-3}	0.0135
461_dom	0.30(0.10)	1.92×10^{-3}	0.0091
861_dom	−0.45(0.10)	1.17×10^{-5}	0.0209
Parameter(s)		$a = -0.01, b = 0.5$	
μ		−0.0600	
σ_0^2		1.5260	
$\hat{h}^2$		0.3976	

[a]add: additive effect; dom: dominance effect. Total number of effects is 10, all with a p-value ≤ 0.01.
[b]The estimated marker effect is denoted by $\hat{\beta}$ and the standard deviation is denoted by $s_{\hat{\beta}}$.
[c]P-value is obtained via t-test.
[d]Phenotypic variation explained.
doi:10.1371/journal.pone.0087330.t002

Table 3. Estimated QTL effects from the full model for the number of grains per panicle.

Loci(i, j)[a]	$\hat{\beta}(s_{\hat{\beta}})$[b]	p-value[c]	$\hat{h}_j^2$[d]
(436$_{add}$, 436$_{add}$)	6.79(0.98)	1.58×10^{-11}	0.0846
(10$_{dom}$, 50$_{add}$)	-8.74(1.41)	1.05×10^{-9}	0.0695
(875$_{add}$, 1156$_{dom}$)	7.15(1.44)	6.37×10^{-7}	0.0427
(595$_{dom}$, 1004$_{add}$)	-12.78(1.88)	3.34×10^{-11}	0.0853
(381$_{dom}$, 1057$_{add}$)	-8.82(1.33)	9.40×10^{-11}	0.0801
Parameter(s)		$a = 0.05$, $b = 0.1$	
μ		-0.3228	
σ_0^2		156.7998	
$\hat{h}^2$		0.4651	

[a]add: additive effect; dom: dominance effect. If i equals j, then it is a main effect, otherwise, it is an interaction between locus i and locus j. Total number of effects is 5, all with a p-value ≤ 0.01.
[b]The estimated marker effect is denoted by $\hat{\beta}$ and the standard deviation is denoted by $s_{\hat{\beta}}$.
[c]P-value is obtained via t-test.
[d]Phenotypic variation explained.
doi:10.1371/journal.pone.0087330.t003

We did a literature survey and found 99 genes with known genomic locations that had experimental evidence showing that they were related to rice yield and yield component traits. For each of the 103 QTLs, we identified genes from 99 experimentally investigated genes that were within 20 centi-Morgan (cM) distance and associated such genes with the QTL. In total, we found 58 genes for 103 QTLs. For the ease of presentation, we organized QTLs within 20 cM distance into a group, which resulted in 51 groups for 103 QTLs. These 51 QTL groups and associated genes are listed in Table S2 in File S1. It is seen that 36 groups of QTLs have at least one associated gene and the distances between QTLs and their associated genes are relatively small (median distance 5.37 cM). Moreover, 21 QTLs involved in 19 of 54 effects locate within 2 cM distance to at least one gene influencing rice yield. The interaction network of the 103 QTLs and their associated genes are visualized in Figure 1.

The three-step CV for the main effect model identified the optimal pair of parameters as $(a, b) = (-0.01, 0.5)$ (Table S1 in File S1), with which eight additive and two dominance effects involving ten QTLs were identified with a p-value ≤ 0.01 (Table 2). The ten effects totally explained 39.76% of the phenotypic variance, and nine of them had genes related to yield within 20 cM distance (median distance 9.29 cM) (Table S3 in File S1). Seven QTLs were identical to QTLs or within the QTL group identified from the full model (Bins 228, 353, 757, 861, 908, 994, 1363), and the other three (Bins 3, 461 and 818) were close to QTLs identified by
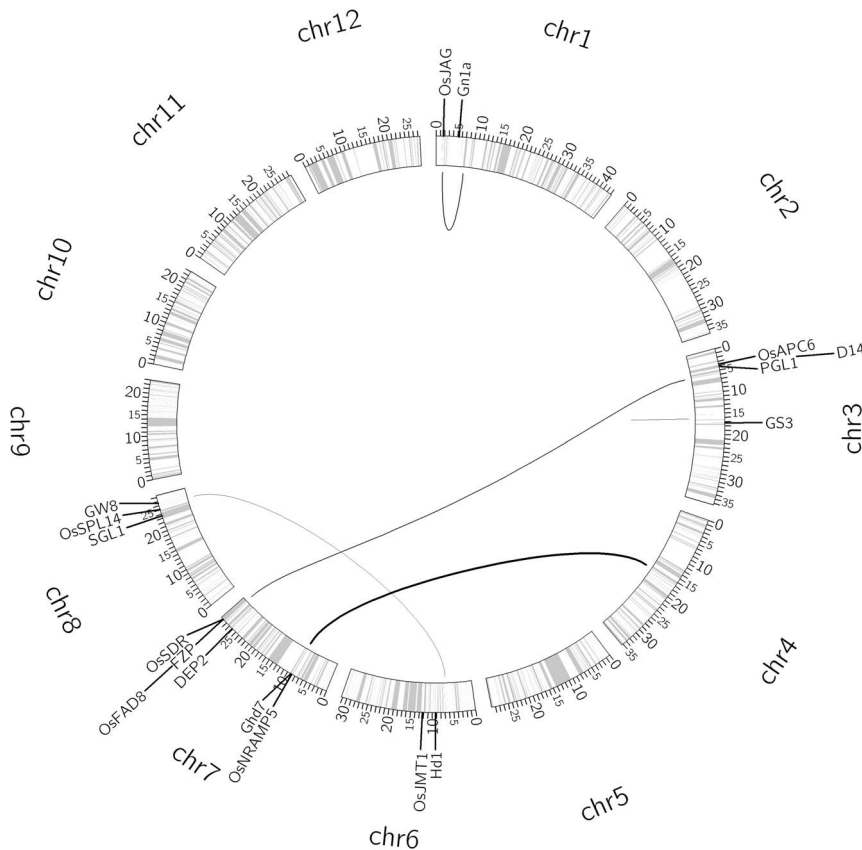
Figure 2. Interaction network of nine QTLs for the number of grains per panicle. The circle shows the bin map and columns indicate position of the makers (ticks in million base pairs). The thickness of a link is proportional to the strength of the interaction effect. A short straight line indicates a main effect. Molecularly characterized genes related to yield are also labeled in the appropriate positions of the genome.
doi:10.1371/journal.pone.0087330.g002

the full model. Specifically, Bin3 was 3.97 cM away from Bin7 identified from the full model; Bin461 was 8.29 cM away from Bin456 identified from the full model; and Bin818 was 6.15 cM away from Bin810 identified from the full model. Comparing the results obtained from the two models, we see that the full model identified more QTLs, which included all those identified by the main effect model, and explained a much larger percentage of the phenotypic variance.

QTL mapping for the number of grains per panicle

The CV analysis identified the optimal pair of parameters $(a, b) = (0.05, 0.1)$ for the full QTL model for the number of grains per panicle (Table S4 in File S1), with which EBlasso identified five nonzero effects. All of these nonzero effects were significant at a p-value ≤ 0.01 (Table 3), including one main additive effect and four additive × dominance interactions. The five effects involved nine QTLs, and explained 46.51% of the overall phenotypic variance. Eight of the nine QTLs have experimentally verified genes related to rice yield within 20 cM distance (median distance 4.86 cM) (Table S5 in File S1). Moreover, four of these QTLs involved in four effects locate within 2 cM distance to at least one yield-related gene. The interaction network of the nine QTLs and their associated genes are depicted in Figure 2.

The same three-step CV for the main effect model identified the optimal pair of parameters $(a, b) = (-0.4, 0.5)$ (Table S4 in File S1), with which five additive effects were identified, all having a p-value ≤ 0.01 (Table 4). The five QTLs (Bins 43, 436, 877, 1006, 1057) totally explained 41.48% of the phenotypic variance, and all had molecularly characterized genes related to rice yield within 19 cM distance (median distance 1.59 cM) (Table S6 in File S1). All five QTLs were identical or very close to the QTLs identified from the full model. Specifically, Bin436 and Bin1057 were identified in both models; Bin43 is 3.40 cM away from Bin50 identified from the full model; Bin877 is 0.47 cM away from Bin875 identified from the full model; and Bin1006 is 0.72 cM away from Bin1004 identified from the full model. Comparing the results obtained from the two models, we observed that although both models identified five effects, the full model identified four more QTLs and explained a slightly larger percentage of phenotypic variance. Moreover, the main effect model identified five additive effects, but the full model identified QTLs with both additive and dominance effects.

QTL mapping for grain weight

The CV analysis determined the optimal $(a, b) = (1, 1)$ (Table S7 in File S1) for the full QTL model for grain weights. Using the optimal a and b, EBlasso yields a QTL model including 89 nonzero effects, among which 28 effects were identified as significant at a p-value ≤ 0.01 (Table 5). Among them, one was a main additive effect, 10 were additive × additive, 15 were additive ×dominance, and two were dominance × dominance interactions. The 28 effects involved 52 QTLs, and explained 93.79% of the phenotypic variance. QTLs with a distance $\leq$ 20 cM were placed into a group, resulting in 32 groups, and 26 of the 32 QTL groups had at least one gene within 20 cM distance (median distance 5.06 cM) (Table S8 in File S1). Moreover, 15 QTLs involved in 12 of 28 effects locate within 2 cM distance to at least one yield-related gene. The interaction network of the 52 QTLs and their associated genes are shown in Figure 3.

The CV analysis for the main effect model identified the optimal pair of parameters $(a, b) = (1, 1)$ (Table S7 in File S1), with which 26 QTLs (19 additive and 7 dominance effects) were identified with a p-value ≤ 0.01 (Table 6). The 26 QTLs totally explained 84.24% of the overall phenotypic variance, and 23 of

Table 4. Estimated QTL effects from the main effect model for the number of grains per panicle.

locus[a]	$\hat{\beta}(s_{\hat{\beta}})$[b]	p-value[c]	$\hat{h}_j^2$[d]
43 _add	−5.15(1.08)	1.38×10^{-6}	0.0454
436 _add	8.06(1.03)	6.02×10^{-14}	0.1190
877 _add	3.64(1.05)	3.21×10^{-4}	0.0225
1006 _add	−8.00(1.15)	1.48×10^{-11}	0.1036
1057 _add	−7.51(1.11)	3.55×10^{-11}	0.0973
Parameter(s)		$a = -0.4, b = 0.5$	
μ		−0.6700	
σ_0^2		171.21	
$\hat{h}^2$		0.4148	

[a]add: additive effect; dom: dominance effect. Total number of effects is five, all with a p-value ≤ 0.01.
[b]The estimated marker effect is denoted by $\hat{\beta}$ and the standard deviation is denoted by $s_{\hat{\beta}}$.
[c]P-value is obtained via t-test.
[d]Phenotypic variation explained.
doi:10.1371/journal.pone.0087330.t004

them had molecularly characterized genes related to rice yield within 16 cM distance (median distance 4.08 cM) (Table S9 in File S1). Twenty three of the 26 QTLs were identical to or within a QTL group identified from the full model, but three QTLs (Bins 228, 843, and 894) do not correspond to any QTLs identified from the full model within 20 cM distance. Again, the full model identified more QTLs than the main effect model and the QTLs detected by the full model explained more phenotypic variance than those detected by the main effect model.

QTL mapping for yield per plant

The CV analysis determined the optimal pair of parameters $(a, b) = (1, 1)$ for the full QTL model for rice yield (Table S10 in File S1). Using the optimal values of (a, b), EBlasso yielded four nonzero effects, all were significant at a p-value ≤ 0.01: one main additive effect, one additive × additive interaction, one additive ×dominance interaction, and one dominance × dominance interaction (see Table 7). The four effects involved seven QTLs and explained 34.01% of the overall phenotypic variance. Five out of the seven QTLs have an experimentally verified gene within 15 cM distance (median distance 2.21 cM) (Table S11 in File S1). Moreover, two QTLs involved in two of four effects locate within 2 cM distance to at least one yield-related gene. The interaction network of the seven QTLs and their associated genes are described in Figure 4.

The optimal pair of parameters determined by the CV analysis for the main effect model was $(a, b) = (-0.5, 0.1)$ (Table S10 in File S1), with which four QTLs with a p-value ≤ 0.01 were identified (Table 8). The four QTL effects explained 23.79% of the phenotypic variance, and all had at least one gene within 17 cM distance (median distance 7.82 cM) (Table S12 in File S1). Two of the four QTLs (Bin1014, Bin1057) were identical to the QTLs identified from the full model, but the other two QTLs do not correspond to any QTL identified from the full model within 20 cM distance. Overall, although the full model did not detect all QTLs identified by the main effect model, it still detected more QTLs and explained more phenotypic variance.

Table 5. Estimated QTL effects from the full model for grain weight.

Loci(i, j)[a]	$\hat{\beta}(s_{\hat{\beta}})$[b]	p-value[c]	$\hat{h}_j^{2\,d}$
(729_{add}, 729_{add})	1.02(0.07)	$<10^{-15}$	0.1548
(37_{add}, 547_{dom})	0.71(0.09)	1.29×10^{-14}	0.0428
(67_{add}, 772_{add})	-0.25(0.08)	6.56×10^{-4}	0.0047
(96_{add}, 1117_{dom})	0.21(0.08)	2.95×10^{-3}	0.0035
(119_{add}, 987_{add})	0.18(0.07)	6.73×10^{-3}	0.0024
(151_{add}, 1262_{add})	-0.15(0.07)	9.79×10^{-3}	0.0018
(71_{dom}, 184_{add})	-0.67(0.09)	1.55×10^{-13}	0.0374
(210_{add}, 1400_{add})	0.19(0.08)	7.79×10^{-3}	0.0025
(329_{add}, 727_{dom})	0.22(0.09)	4.63×10^{-3}	0.0040
(310_{dom}, 419_{add})	-0.21(0.05)	1.86×10^{-5}	0.0043
(431_{add}, 1111_{add})	0.35(0.08)	1.01×10^{-5}	0.0107
(71_{dom}, 500_{add})	-0.76(0.08)	$<10^{-15}$	0.0493
(583_{add}, 1578_{dom})	0.35(0.08)	9.50×10^{-6}	0.0092
(107_{dom}, 700_{add})	0.19(0.07)	4.80×10^{-3}	0.0035
(708_{dom}, 714_{add})	-1.15(0.32)	1.79×10^{-4}	0.0076
(818_{add}, 1100_{add})	0.26(0.08)	3.93×10^{-4}	0.0053
(916_{add}, 1026_{add})	0.15(0.06)	8.50×10^{-3}	0.0019
(472_{dom}, 920_{add})	-0.27(0.09)	1.65×10^{-3}	0.0058
(18_{dom}, 955_{add})	-0.20(0.08)	3.77×10^{-3}	0.0033
(971_{add}, 1461_{add})	0.27(0.08)	4.75×10^{-4}	0.0075
(620_{dom}, 1011_{add})	-0.67(0.09)	3.31×10^{-12}	0.0336
(1035_{add}, 1224_{add})	0.30(0.07)	3.27×10^{-5}	0.0081
(1093_{add}, 1407_{dom})	0.44(0.08)	2.13×10^{-7}	0.0148
(1167_{dom}, 1168_{add})	-0.47(0.16)	1.37×10^{-3}	0.0051
(119_{dom}, 1375_{add})	0.61(0.09)	7.83×10^{-11}	0.0289
(1397_{add}, 1505_{add})	0.41(0.09)	1.95×10^{-6}	0.0119
(247_{dom}, 1505_{dom})	-0.23(0.08)	1.22×10^{-3}	0.0032
(647_{dom}, 796_{dom})	0.26(0.08)	4.29×10^{-4}	0.0044
Parameter(s)		$a=1, b=1$	
μ		-0.0661	
σ_0^2		0.5317	
$\hat{h}^2$		0.9379	

[a]add: additive effect; dom: dominance effect. If i equals j, then it is a main effect, otherwise, it is an interaction between locus i and locus j. Total number of effects is 90, only 28 effects with a p-value ≤0.01 are listed in this table.
[b]The estimated marker effect is denoted by $\hat{\beta}$ and the standard deviation is denoted by $s_{\hat{\beta}}$.
[c]P-value is obtained via t-test.
[d]Phenotypic variation explained.
doi:10.1371/journal.pone.0087330.t005

Effect types and pleiotropic genes

Among the five types of effects (main additive, main dominance effects, additive × additive, additive ×dominance, and dominance × dominance interactions) considered in the EBlasso full models for four traits, no main dominance effects was detected, but several dominance × dominance interactions (one for rice yield, three for the number of panicles per plant, and two for grain weight) were identified. Many additive ×dominance interaction effects were identified, including one for rice yield, 32 for the number of panicles per plant, four for the number of grains per panicle, and 15 for grain weight. Phenotypic variance explained by a single

effect is relatively small for all traits (Tables 1, 3, 5 and 7). For example, the largest effect has $\hat{h}^2 = 7.82\%$ ($908_{dominance}$ $\times994_{additive}$) for the number of panicles per plant, 8.53% ($595_{dominance}\times1004_{additive}$) for the number of grains per panicle, 15.48% ($729_{additive}$) for grain weight, and 6.08% ($1057_{additive}\times$ $1144_{dominance}$) for yield per plant. Each main effect detected by the main effect model also explained a small percentage of the total phenotypic variance.

Many molecularly characterized genes related to yield are known to play pleiotropic roles in regulating grain productivity [31]. Without surprise, a number of such genes coincide with or close to the QTLs that were identified by our EBlasso for multiple traits, although they did not necessarily have pleiotropic effects. For example, gene *Ghd7*, *OsNRAMP5* and *DEP2* are close to several QTLs common for the four phenotypes, *qSW5/GW5*, *OsEF3* and *LOG* are near the QTLs for three phenotypes except the number of grains per panicle, and *Gn1a*, *OsJAG*, *GS3*, *OsJMT1*, *OsSPL14*, *GW8/OsSPL16*, *SGL1* are associated with QTLs for three phenotypes except yield per plant. Besides *Ghd7*, *OsNRAMP5* and *DEP2*, gene *FZP*, *OsSDR*, and *OsFAD8* was near QTLs for both yield per plant and the number of grains per panicle; 14 genes were close to QTLs for both the number of panicles per plant and the number of grains per panicle; and 62 other genes were associated with QTLs for both the number of grains per panicle and grain weight. While the pleiotropic effect of some genes have been reported [32], our QTL mapping results identified a number of genes associated with multiple phenotypes, implying their possible pleiotropic role worthy of further experimental investigation. Moreover, it is also possible that the QTLs we detected may be closely linked to unknown genes, which, if identified, will yield more insight into the molecular basis of phenotypes [1].

Discussion

Due to its small genome and close relatedness with other grass crops, rice has served as a model plant for investigating genetic factors underlying crop productivity [33,34]. To date, more than 600 rice genes have been experimentally cloned with related traits including yield, biotic and abiotic stresses, grain quality, plant architecture, fertility, etc. [2]. However, there is still a knowledge gap regarding the molecular basis of yield-related biological processes [1], suggesting the importance of systematic tools that can enable to understand functional role of genes [2,35]. In this study, we employed a multiple QTL model that included all additive and dominance main effects of 1,619 markers, and all their pair-wise interactions with a total of more than 5 million possible effects, and then applied our EBlasso algorithm to identify QTLs for four agronomic related traits of rice, including yield, the number of panicles per plant, the number of grains per panicle and grain weight. Our QTL mapping revealed a number of QTLs for four traits, most of which are involved in digenic interactions. Moreover, most of these QTLs have at least one experimentally cloned gene within 20 cM distance.

The same set of markers in the recombinant inbred line (RIL) population where the "immortalized F_2" (IMF$_2$) was derived from were used for QTL mapping, via a composite interval mapping method with a scan window size of five markers [12]. Upon development of the IMF$_2$ population, this dataset was obtained and the ANOVA method was applied to each pair of markers to identify both main and digenic interaction effects from 5,242,322 possible effects [24]. The composite interval mapping identified zero, three (Bin40, Bin446 and Bin1006), seven (Bins 49, 171, 439, 729, 928, 1008, and 1266), and one (Bin1007) QTLs for the
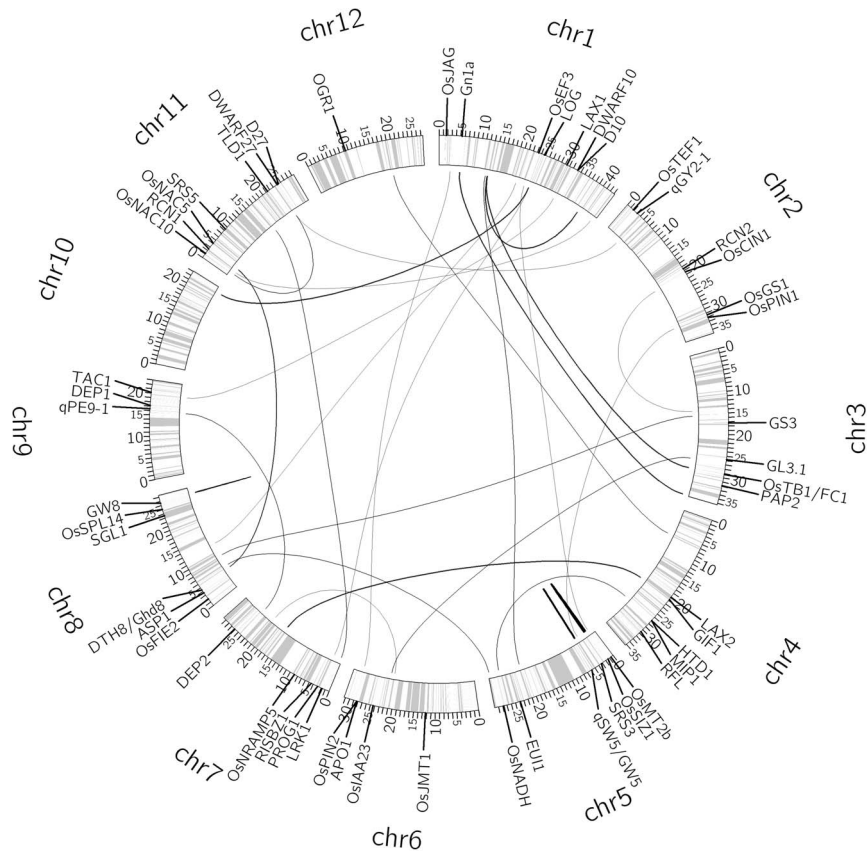
Figure 3. Interaction network of 52 QTLs for grain weight. The circle shows the bin map and columns indicate position of the makers (ticks in million base pairs). The thickness of a link is proportional to the strength of the interaction effect. A short straight line indicates a main effect. Molecularly characterized genes related to yield are also labeled in the appropriate positions of the genome.
doi:10.1371/journal.pone.0087330.g003

number of panicles per plant, the number of grains per panicle, grain weight and yield per plant, respectively. The ANOVA method detected thousands (1432, 2696, 3524 and 2251) of digenic interactions between two bins with a p-value ≤ 0.001; and after those digenic interactions involving adjacent bins were merged, 115, 189, 238, and 204 effects were reported, respectively [24]. In contrast, our EBlasso method identified a reasonable number of effects and QTLs for each trait, and 35%–80% of identified effects for four traits involve at least one QTL that locates within 2 cM distance to at least one gene related to crop yield, which corroborates the reliability of the identified effects.

The list of genes associated with the identified QTLs provides insight into rice yield with respect to yield component traits. First, the number of panicles depends on plant's ability of producing tillers, which is under genetic, developmental and environmental influence. While previous composite interval mapping did not identify any significant effect with the same set of markers in an RIL population [12], we have identified a set of QTLs that have nearby genes known to regulate plant tillering. For example, among genes in Table S2 in File S1, *MOC1/SPA* is the first gene characterized for rice tillering; it initiates axillary buds that grow into lateral braches [36]. *OsTB1/FC1* has been identified as an important gene that negatively regulates lateral branching in rice [37]. *OsSPL14* is a highly expressed gene in the shoot apex and primordial of primary and secondary branches, which promotes panicle branching while reducing tiller number [38]. Through

gene mutations, *D3*, *D10*, *D14*, *D17/HTD1*, and *D27* were found to affect tiller initiation and/or outgrowth [37]. Secondly, the number of grains per panicle is another important trait determining crop yield. While composite interval mapping identified three QTLs (Bin40, Bin446 and Bin1006) close to genes *Gn1a*, *GS3*, *OsNRAMP5* and *Ghd7*, our EBlasso also identified these genes in addition to other 13 genes. Among them, *FZP* is known to control spikelet meristem identity [39], *Ghd7* is a pleiotropic gene affecting grain number, plant height and heading date [40], *GW8/OsSPL16*, *PGL*, and *DEP2* all are known to be essential in regulating cell proliferation or elongation [41–43]. Thirdly, composite interval mapping detected seven QTLs (Bins 49, 171, 439, 729, 928, 1008, and 1266) for grain weight, with nearby genes *Gn1a*, *LAX1*, *GS3*, *GS5*, *qSW5/GW5*, *OsJMT1*, *OsIAA23*, *Ghd7*, *OsNRAMP5*, *TAC1*, *LGD1* and *SG1*. In addition to these genes, our EBlasso identified many other genes with known effects in controlling grain weight. For example, *GIF1* is a gene encoding a cell-wall invertase required for carbon partitioning during early grain filling, and overexpression of *GIF1* leads to larger and heavier grain weight [44]. Genes *SRS3* and *SRS5* have been found to regulate seed cell elongation [45,46]. Over-expression of *LRK1* gene results in enhanced cellular proliferation and increased grain weight [47]. Finally, yield per plant is the most complex trait and a small number of effects were identified compared with its component traits. While composite interval mapping identified only one QTL (Bin1007) with nearby gene *Ghd7* and *OsNRAMP5*,

Table 6. Estimated QTL effects from the main effect model for grain weight.

locus[a]	$\hat{\beta}(s_{\hat{\beta}})^b$	p-value[c]	$\hat{h}_j^{2\,d}$
37_add	0.40(0.08)	8.22×10^{-7}	0.0262
50_add	0.21(0.08)	3.14×10^{-3}	0.0072
151_add	−0.18(0.07)	2.77×10^{-3}	0.0052
173_add	−0.49(0.07)	6.46×10^{-11}	0.0418
199_add	0.23(0.06)	1.51×10^{-4}	0.0077
332_add	0.30(0.06)	1.26×10^{-6}	0.0150
440_add	−0.98(0.06)	$<10^{-15}$	0.1670
498_add	−0.35(0.06)	5.18×10^{-8}	0.0188
710_add	0.17(0.06)	2.13×10^{-3}	0.0047
729_add	0.81(0.07)	$<10^{-15}$	0.0968
894_add	−0.18(0.06)	8.21×10^{-4}	0.0053
936_add	−0.44(0.07)	3.11×10^{-10}	0.0291
1008_add	−0.34(0.06)	6.63×10^{-8}	0.0177
1110_add	0.18(0.05)	4.85×10^{-4}	0.0055
1176_add	−0.26(0.06)	8.77×10^{-6}	0.0108
1251_add	0.37(0.07)	4.64×10^{-8}	0.0171
1374_add	0.33(0.06)	2.55×10^{-7}	0.0167
1442_add	−0.22(0.06)	6.37×10^{-5}	0.0079
1565_add	0.20(0.06)	2.33×10^{-4}	0.0063
38_dom	0.29(0.07)	3.18×10^{-5}	0.0066
228_dom	−0.20(0.07)	1.60×10^{-3}	0.0032
312_dom	0.17(0.06)	3.20×10^{-3}	0.0024
441_dom	−0.23(0.07)	7.95×10^{-4}	0.0043
547_dom	0.15(0.06)	6.68×10^{-3}	0.0019
843_dom	0.16(0.06)	5.41×10^{-3}	0.0021
1506_dom	−0.26(0.07)	1.89×10^{-4}	0.0054
Parameter(s)		$a = 1, b = 1$	
μ		0.1000	
σ_0^2		0.5224	
$\hat{h}^2$		0.8424	

[a]add: additive effect; dom: dominance effect. Total number of effects is 38, only 30 effects with a p-value ≤0.01 are listed in this table.
[b]The estimated marker effect is denoted by $\hat{\beta}$ and the standard deviation is denoted by $s_{\hat{\beta}}$.
[c]P-value is obtained via t-test.
[d]Phenotypic variation explained.
doi:10.1371/journal.pone.0087330.t006

our EBlasso identified this QTL and six other QTLs, four of which have cloned gene within 15 cM distance (Table S11 in File S1).

In conclusion, taking advantage of the powerful EBlasso model for simultaneously accounting for more than 5 million possible effects, we identified a number of QTLs for four traits of the elite rice hybrid Shanyou 63, a vast majority of which are involved in digenic interactions. This set of QTLs not only shed light on the genetic basis of the yield of the rice hybrid, but also provide candidate loci for identification of new genes that may be involved in crop yield.

Materials and Methods

Plant materials and QTLs

The genotype and phenotype data used in this study were obtained from previous studies [12,24]. The mapping plants were created by first crossing between *indica* rice Zhensha 97 and Minghui 63 [7] to produce the elite rice hybrid Shanyou 63 that was the most widely cultivated in China in 1980s –1990s [24]. Then a population of 240 F_9 RILs was derived from single-seed descent of Shanyou 63. Next, an "immortalized F_2" (IMF_2) population consisted of 278 crosses was created by intercrossing RILs for QTL mapping study [7,23]. The crossed population was field tested on the experimental farm of Huazhong Agricultural University in Wuhan, China, in 1999, for traits including yield per plant, the number of panicles per plant, the number of grains per panicle and grain weight.

The RILs were genomic sequenced with an Illumina Genome Analyzer II using the bar-coded multiplexed sequencing approach as described in [25], and 270,820 high quality SNPs were identified. Bin maps were constructed by lumping consecutive SNPs with the same genotype into blocks, masking blocks with less than 250 kb to avoid false double recombinations, and merging recombination bins less than 5 kb, resulting in a map consisting of 1,619 bins without missing data [12]. Genotypes of the IMF_2 crosses were deduced according to genotypes of their RIL parents [24]. The three genotypes in each bin were coded as A and B for each parental homozygote genotype and H for the heterozygote. Using the recombinant bins as QTLs, a 1,625.5 cM genetic linkage map was constructed with about 1.0 cM (230 kb) in length per bin (Figure 1).

Bayesian Lasso linear regression model for multiple QTLs

We employed a Bayesian Lasso (BLasso) multiple linear regression model to infer genotypes and quantitative trait associations. The regression model includes main additive and dominance effects of 1,619 SNP bins and all their pair-wise interactions. Let y_i be the phenotypic value of a quantitative trait of the ith individual in a mapping population. In this study we observed y_i, $i = 1, \cdots, n$, of $n = 278$ individuals and collected them into a vector $\mathbf{y} = [y_1, y_2, \cdots, y_n]^T$. In these n individuals, let

Table 7. Estimated QTL effects from the full model for yield per plant.

Loci(i, j)[a]	$\hat{\beta}(s_{\hat{\beta}})^b$	p-value[c]	$\hat{h}_j^{2\,d}$
(1014_add, 1014_add)	−1.94(0.37)	1.95×10^{-7}	0.0544
(113_add, 1547_add)	−2.81(0.53)	1.43×10^{-7}	0.0552
(1057_add, 1144_dom)	−2.89(0.53)	5.57×10^{-8}	0.0608
(743_dom, 1043_dom)	3.21(0.52)	8.38×10^{-10}	0.0598
Parameter(s)		$a = 1, b = 1$	
μ		−0.7521	
σ_0^2		22.9734	
$\hat{h}^2$		0.3401	

[a]add: additive effect; dom: dominance effect. If i equals j, then it is a main effect, otherwise it is an interaction between locus i and locus j. Total number of effects is 4, all with a p-value ≤0.01.
[b]The estimated marker effect is denoted by $\hat{\beta}$ and the standard deviation is denoted by $s_{\hat{\beta}}$.
[c]P-value is obtained via t-test.
[d]Phenotypic variation explained.
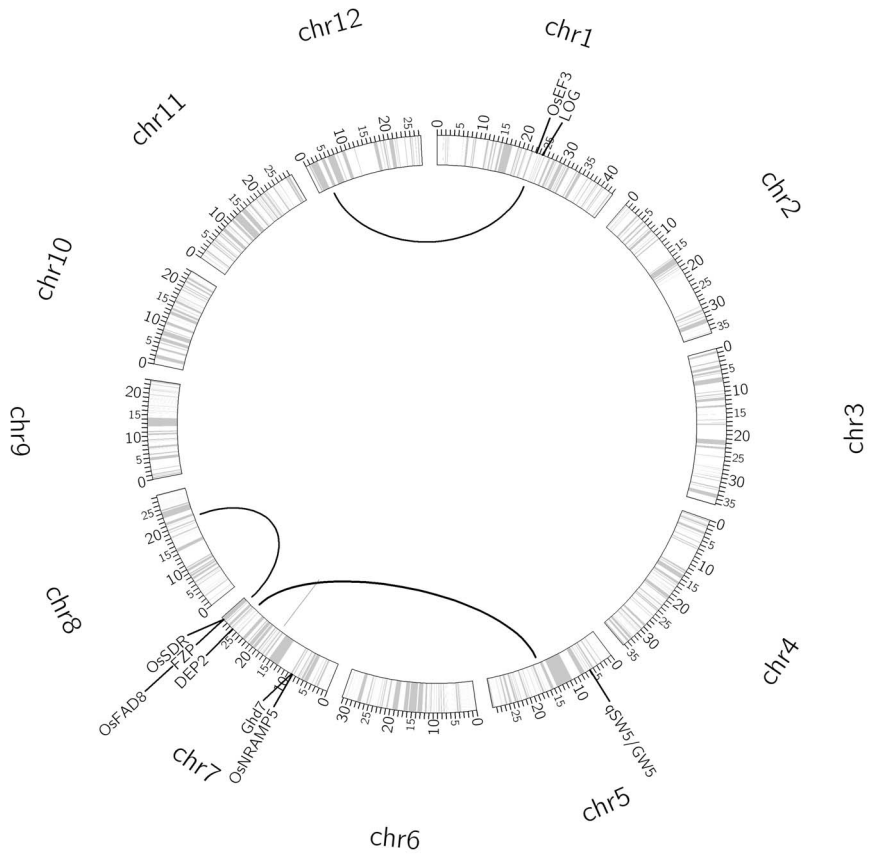doi:10.1371/journal.pone.0087330.t007

Figure 4. Interaction network of seven QTLs for yield per plant. The circle shows the bin map and columns indicate position of the makers (ticks in million base pairs). The thickness of a link is proportional to the strength of the interaction effect. A short straight line indicates a main effect. Molecularly characterized genes related to yield are also labeled in the appropriate positions of the genome.
doi:10.1371/journal.pone.0087330.g004

$m = 1,619$ denote the number of genetic markers genotyped whose main effects include additive and dominance effects. Let the additive and dominance genotypes of marker j of individual i be x_{Aij} and x_{Dij}, respectively, where x_{Aij} takes on values $+1$, 0 and -1, and x_{Dij} takes on values 0, $+1$ and 0, corresponding to genotypes A, H and B, respectively. Let us define $\mathbf{x}_{Ai} = [x_{Ai1}, x_{Ai2}, \cdots, x_{Aim}]^T$ and $\mathbf{x}_{Gi} = [x_{Di1}, x_{Di2}, \cdots, x_{Dim}]^T$. The interactions between any two effects are modeled as element-wise product of the corresponding main effects. Let $\mathbf{x}_{AAi}$, $\mathbf{x}_{ADi}$, $\mathbf{x}_{DAi}$, and $\mathbf{x}_{DDi}$ be $m(m-1)/2 \times 1$ vectors containing $x_{Aij} \cdot x_{Aij'}$, $x_{Aij} \cdot x_{Dij'}, x_{Dij} \cdot x_{Aij'}$, and $x_{Dij} \cdot x_{Dij'}$, respectively, where $j = 1, \cdots, m-1$ and $j' > j$. Then we have the following linear regression model for $\boldsymbol{y}$:

$$\boldsymbol{y} = \mu + \mathbf{X}_A\boldsymbol{\beta}_A + \mathbf{X}_D\boldsymbol{\beta}_D + \mathbf{X}_{AA}\boldsymbol{\beta}_{AA} + \mathbf{X}_{AD}\boldsymbol{\beta}_{AD} + \mathbf{X}_{DA}\boldsymbol{\beta}_{DA} + \mathbf{X}_{DD}\boldsymbol{\beta}_{DD} + \boldsymbol{\varepsilon}, \quad (2)$$

where μ is the population mean, vectors $\boldsymbol{\beta}_A$ and $\boldsymbol{\beta}_D$ represent the main additive and dominance effects of all markers, respectively, and vectors $\boldsymbol{\beta}_{AA}$, $\boldsymbol{\beta}_{AD}$, $\boldsymbol{\beta}_{DA}$ and $\boldsymbol{\beta}_{DD}$ capture the additive × additive, additive × dominance, dominance × additive, and dominance × dominance interactions, respectively. Matrices $\mathbf{X}_A = [\mathbf{x}_{A1}, \mathbf{x}_{A2}, \cdots, \mathbf{x}_{An}]^T$, $\mathbf{X}_D = [\mathbf{x}_{D1}, \mathbf{x}_{D2}, \cdots, \mathbf{x}_{Dn}]^T$, $\mathbf{X}_{AA} = [\mathbf{x}_{AA1}, \mathbf{x}_{AA2}, \cdots, \mathbf{x}_{AAn}]^T$, $\mathbf{X}_{AD} = [\mathbf{x}_{AD1}, \mathbf{x}_{AD2}, \cdots, \mathbf{x}_{ADn}]^T$, $\mathbf{X}_{DA} = [\mathbf{x}_{DA1}, \mathbf{x}_{DA2}, \cdots, \mathbf{x}_{DAn}]^T$, and $\mathbf{X}_{DD} = [\mathbf{x}_{DD1}, \mathbf{x}_{DD2}, \cdots, \mathbf{x}_{DDn}]^T$ are the corresponding design matrices of different effects, and 1ε is the residual error that follows a normal distribution with zero mean and variance $\sigma_0^2\mathbf{I}$.

Given m markers, the size of matrix $\mathbf{X}_A$ or $\mathbf{X}_D$ is $n \times m$, and the size of $\mathbf{X}_{AA}$, $\mathbf{X}_{AD}$, $\mathbf{X}_{DA}$, or $\mathbf{X}_{DD}$ is $n \times q$, where $q = m(m-1)/2 = 1,309,771$. Defining $\boldsymbol{\beta} = [\boldsymbol{\beta}_A^T, \boldsymbol{\beta}_D^T, \boldsymbol{\beta}_{AA}^T, \boldsymbol{\beta}_{AD}^T, \boldsymbol{\beta}_{DA}^T, \boldsymbol{\beta}_{DD}^T]^T$, and $\mathbf{X} = [\mathbf{X}_A, \mathbf{X}_D, \mathbf{X}_{AA}, \mathbf{X}_{AD}, \mathbf{X}_{DA}, \mathbf{X}_{DD}]$, we can write (2) in a more compact form:

$$\boldsymbol{y} = \mu + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (3)$$

The size of matrix $\mathbf{X}$ is $n \times k$, where $k = 2m + 4q = 5,242,322$, and we apparently have $k \gg n$. However, we would expect that most elements of $\boldsymbol{\beta}$ are zeros and thus we have a sparse linear model. The Blasso model employs a three-level hierarchical prior distribution to model the sparsity. At the first level, let $\beta_j, j = 1, 2, \cdots, k$, follows an independent normal distribution with mean zero and unknown variance $\sigma_j^2 : \beta_j \tilde{N}(0, \sigma_j^2)$. At the second level, let $\sigma_j^2, j = 1, 2, \cdots, k$, follows an independent exponential distribution with a common parameter λ: $p(\sigma_j^2) = \lambda\exp(-\lambda\sigma_j^2)$. At the third level, we assign a conjugate Gamma prior $Gamma(a, b)$ with a shape parameter a and an inverse scale parameter b to the parameter λ. Finally, we assign non-informative uniform priors to μ and σ_0^2. The three-level hierarchical model has two hyperparameters (a, b) for adjusting the degree of shrinkage, and cross validation (CV) can be applied to choose appropriate values of these parameters.

The QTL model (2) or equivalently (3) includes all main effects and digenic interactions. We refer to this model as the full model throughout the paper. We also performed QTL mapping with the model $\boldsymbol{y} = \mu + \mathbf{X}_A\boldsymbol{\beta}_A + \mathbf{X}_D\boldsymbol{\beta}_D + \boldsymbol{\varepsilon}$, which is referred to as the main effect model, since it includes only the main effects.

Model inference and cross validation

The Blasso model can be inferred efficiently with the empirical Blasso (EBlasso) algorithm [21]. The EBlasso algorithm employs a coordinate ascent method to find $\hat{\sigma}_j^2$, the estimate of σ_j^2, $j = 0 \ldots, k$, that maximizes the likelihood function of σ_j^2, $j = 0, \ldots, k$. In the iterative process, many σ_j^2 or equivalent β_j are shrunk to zero. The coordinate ascent method along with other algorithmic techniques makes the EBlasso algorithm very efficient. Our previous studies demonstrated that EBlasso outperformed several other multiple QTL mapping methods including the empirical Bayes method [26], the Bayesian hierarchical generalized linear models (BhGLM) [27], HyperLasso [28], and Lasso [29]. Detailed description of the EBlasso algorithm can be found in [21,22] and an efficient C program with the R interface [30] implementing the EBlasso algorithm is available.

The optimal values of two hyperparameters (a, b) of the EBlasso algorithm were obtained with five-fold CV in three steps to minimize the prediction error (PE) calculated from $PE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, where $\hat{y}_i, i = 1, \cdots, n$, is the estimated phenotypic value. In the first step, $a = b = 0.001, 0.01, 0.1, 1$ were examined and a pair (a_1, b_1) corresponding to the smallest PE was obtained. In the second step, b was fixed at b_1 and a was chosen from the set $[-0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, -0.01, 0.01, 0.05, 0.1, 0.5, 1]$, which yielded a value a_2 corresponding to the smallest PE. In the third step, $a = a_2$ was fixed and b varied from 0.01 to 10 with a step size of one for $b > 1$ and a step size of one on the logarithmic scale for $b < 1$. Note that when fixing one of the two parameters, the degree of shrinkage is a monotonic function of the other parameter [21,22]. Therefore, in the second and third steps, the selection did not go through the full path but stopped if the current PE was one standard error larger than the minimum PE in previous steps.

Statistical significance test

One advantage of the EBlasso algorithm relative to Lasso [29] is that it not only outputs a $k' \times 1$ ($k' \ll k$) vector $\hat{\boldsymbol{\beta}}$ as an estimate of nonzero elements of $\boldsymbol{\beta}$, but also gives an estimate of the covariance of $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\Sigma}}$. Letting $\hat{\boldsymbol{\Sigma}}_{jj}$ be the jth diagonal element of $\hat{\boldsymbol{\Sigma}}$, we can use the t-statistics $\hat{\beta}_j / \hat{\Sigma}_{jj}^{1/2}$ to test if $\hat{\beta}_j \neq 0$ at a certain significance level.

Supporting Information

File S1 Tables S1–S12. Table S1. Cross-validation for determining hyperparameters (a, b) used in QTL mapping for

Table 8. Estimated QTL effects from the main effect model for yield per plant.

locus[a]	$\hat{\beta}(s_{\hat{\beta}})$[b]	p-value[c]	$\hat{h}_j^2$[d]
181$_{add}$	−1.87(0.42)	6.24×10^{-6}	0.0518
1014$_{add}$	−2.27(0.43)	9.58×10^{-8}	0.0743
1057$_{add}$	−2.41(0.43)	2.58×10^{-8}	0.0848
1100$_{add}$	1.17(0.39)	1.35×10^{-3}	0.0205
Parameter(s)		$a = -0.5, b = 0.1$	
μ		−0.1500	
σ_0^2		26.2580	
$\hat{h}^2$		0.2379	

[a]add: additive effect; dom: dominance effect. Total number of effects is 4, all with a p-value ≤0.01.
[b]The estimated marker effect is denoted by $\hat{\beta}$ and the standard deviation is denoted by $s_{\hat{\beta}}$.
[c]P-value is obtained via t-test.
[d]Phenotypic variation explained.
doi:10.1371/journal.pone.0087330.t008

the number of panicles per plant. **Table S2.** Experimentally investigated genes near QTLs for the number of panicles per plant identified with the full model. **Table S3.** Experimentally investigated genes near QTLs for the number of panicles per plant identified with the main effect model. **Table S4.** Cross-validation for determining hyperparameters (a, b) used in QTL mapping for the number of grains per panicle. **Table S5.** Experimentally investigated genes near QTLs for the number of grains per panicle identified with the full model. **Table S6.** Experimentally investigated genes near QTLs for the number of grains per panicle identified with the main effect model. **Table S7.** Cross-validation for determining hyperparameters (a, b) used in QTL mapping for grain weight. **Table S8.** Experimentally investigated genes near QTLs for grain weight identified with the full model. **Table S9.** Experimentally investigated genes near QTLs for grain weight identified with the main effect model. **Table S10.** Cross-validation for determining hyperparameters (a, b) used in QTL mapping for yield per plant. **Table S11.** Experimentally investigated genes near QTLs for yield per plant identified with the full model. **Table S12.** Experimentally investigated genes near QTLs for yield per plant identified with the main effect model.
(DOC)

Author Contributions

Conceived and designed the experiments: XC SX. Performed the experiments: AH. Analyzed the data: AH XC SX. Wrote the paper: AH XC SX.

References

1. Xing Y, Zhang Q (2010) Genetic and molecular bases of rice yield. Annu Rev Plant Biol 61: 421–442.
2. Jiang Y, Cai Z, Xie W, Long T, Yu H, et al. (2012) Rice functional genomics research: Progress and implications for crop genetic improvement. Biotechnol Adv 30: 1059–1070.
3. Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327: 818–822.
4. Ikeda M, Miura K, Aya K, Kitano H, Matsuoka M (2013) Genes offering the potential for designing yield-related traits in rice. Curr Opin Plant Biol 16: 213–220.
5. Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185–199.
6. Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20: 176–183.
7. Hua JP, Xing YZ, Xu CG, Sun XL, Yu SB, et al. (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 162: 1885–1895.
8. Li J, Yu S, Xu C, Tan Y, Gao Y, et al. (2000) Analyzing quantitative trait loci for yield using a vegetatively replicated F2 population from a cross between the parents of an elite rice hybrid. Theor Appl Genet 101: 248–254.

9. Xing Y, Tan Y, Hua J, Sun X, Xu C, et al. (2002) Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor Appl Genet 105: 248–257.

10. Tan Y-F, Xing Y-Z, Li J-X, Yu S-B, Xu C-G, et al. (2000) Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. Theor Appl Genet 101: 823–829.

11. Lian X, Xing Y, Yan H, Xu C, Li X, et al. (2005) QTLs for low nitrogen tolerance at seedling stage identified using a recombinant inbred line population derived from an elite rice hybrid. Theor Appl Genet 112: 85–96.

12. Yu H, Xie W, Wang J, Xing Y, Xu C, et al. (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. PLoS ONE 6: e17595.

13. Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136: 1457–1468.

14. Song X-J, Ashikari M (2008) Toward an optimum return from crop plants. Rice 1: 135–143.

15. Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48: 1649–1664.

16. Xu S (2003) Estimating polygenic effects using markers of the entire genome. Genetics 163: 789–801.

17. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

18. Zhou X, Carbonetto P, Stephens M (2013) Polygenic modeling with Bayesian sparse linear mixed models. PLoS Genet 9: e1003264.

19. Iwata H, Ebana K, Uga Y, Hayashi T, Jannink J-L (2010) Genome-wide association study of grain shape variation among Oryza sativa L. germplasms based on elliptic Fourier analysis. Mol Breed 25: 203–215.

20. de los Campos G, Pérez P, Vazquez AI, Crossa J (2013) Genome-enabled prediction using the BLR (Bayesian Linear Regression) R-package. Genome-Wide Association Studies and Genomic Prediction. New York: Springer. pp. 299–320.

21. Cai X, Huang A, Xu S (2011) Fast empirical Bayesian LASSO for multiple quantitative trait locus mapping. BMC Bioinformatics 12: 211.

22. Huang A, Xu S, Cai X (2013) Empirical Bayesian LASSO-logistic regression for multiple binary trait locus mapping. BMC Genet 14: 5.

23. Hua J, Xing Y, Wu W, Xu C, Sun X, et al. (2003) Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 100: 2574–2579.

24. Zhou G, Chen Y, Yao W, Zhang C, Xie W, et al. (2012) Genetic composition of yield heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 109: 15847–15852.

25. Xie W, Feng Q, Yu H, Huang X, Zhao Q, et al. (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci USA 107: 10578–10583.

26. Xu S (2007) An empirical Bayes method for estimating epistatic effects of quantitative trait loci. Biometrics 63: 513–521.

27. Yi N, Banerjee S (2009) Hierachical generalized linear models for multiple quantitative trait locus mapping. Genetics 181: 1101–1133.

28. Hoggart CJ, Whittaker JC, De Iorio M, Balding DJ (2008) Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. PLoS Genet 4: e1000130.

29. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Roy Stat Soc B Met 58: 267–288.

30. R Development Core Team (2012) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

31. Miura K, Ashikari M, Matsuoka M (2011) The role of QTLs in the breeding of high-yielding rice. Trends Plant Sci 16: 319–326.

32. Yan W-H, Wang P, Chen H-X, Zhou H-J, Li Q-P, et al. (2011) A major QTL, Ghd8, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. Mol Plant 4: 319–330.

33. Yu J, Hu S, Wang J, Wong GK-S, Li S, et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296: 79–92.

34. Goff SA, Ricke D, Lan T-H, Presting G, Wang R, et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296: 92–100.

35. Zhang Q, Li J, Xue Y, Han B, Deng XW (2008) Rice 2020: A call for an international coordinated effort in rice functional genomics. Mol Plant 1: 715–719.

36. Li X, Qian Q, Fu Z, Wang Y, Xiong G, et al. (2003) Control of tillering in rice. Nature 422: 618–621.

37. Minakuchi K, Kameoka H, Yasuno N, Umehara M, Luo L, et al. (2010) FINE CULM1 (FC1) works downstream of strigolactones to inhibit the outgrowth of axillary buds in rice. Plant Cell Physiol 51: 1127–1135.

38. Jiao Y, Wang Y, Xue D, Wang J, Yan M, et al. (2010) Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. Nat Genet 42: 541–544.

39. Chuck G, Muszynski M, Kellogg E, Hake S, Schmidt RJ (2002) The control of spikelet meristem identity by the branched silkless1 gene in maize. Science 298: 1238–1241.

40. Xue W, Xing Y, Weng X, Zhao Y, Tang W, et al. (2008) Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat Genet 40: 761–767.

41. Li F, Liu W, Tang J, Chen J, Tong H, et al. (2010) Rice DENSE AND ERECT PANICLE 2 is essential for determining panicle outgrowth and elongation. Cell Res 20: 838–849.

42. Wang S, Wu K, Yuan Q, Liu X, Liu Z, et al. (2012) Control of grain size, shape and quality by OsSPL16 in rice. Nat Genet 44: 950–954.

43. Heang D, Sassa H (2012) Antagonistic actions of HLH/bHLH proteins are involved in grain length and weight in rice. PLoS ONE 7: e31325.

44. Wang E, Xu X, Zhang L, Zhang H, Lin L, et al. (2010) Duplication and independent selection of cell-wall invertase genes GIF1 and OsCIN1 during rice evolution and domestication. BMC Evol Biol 10: 108.

45. Kitagawa K, Kurinami S, Oki K, Abe Y, Ando T, et al. (2010) A novel kinesin 13 protein regulating rice seed length. Plant Cell Physiol 51: 1315–1329.

46. Segami S, Kono I, Ando T, Yano M, Kitano H, et al. (2012) Small and round seed 5 gene encodes alpha-tubulin regulating seed cell elongation in rice. Rice 5: 1–10.

47. Zha X, Luo X, Qian X, He G, Yang M, et al. (2009) Over-expression of the rice LRK1 gene improves quantitative yield components. Plant Biotechnol J 7: 611–620.