Identification of Enzyme Activity Quantitative Trait Loci in a *Solanum lycopersicum* x *Solanum pennellii* introgression line population

- Supplemental Information -

Marie-Caroline Steinhauser[1,*], Dirk Steinhauser[1], Yves Gibon[1,2], Marie Bolger[1], Stéphanie Arrivault[1], Bjoern Usadel[1], Dani Zamir[3], Alisdair R. Fernie[1], Mark Stitt[1]

[1] Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany

[2] Present address: INRA Bordeaux, University of Bordeaux, UMR1332 Fruit Biology and Pathology, F-33883 Villenave d'Ornon, France

[3] The Hebrew University of Jerusalem, Faculty of Agriculture, P.O. Box 12, Rehovot 76100 Israel

* Corresponding author.

Corresponding author:

Marie-Caroline Steinhauser

Max Planck Institute of Molecular Plant Physiology,

Am Muehlenberg 1,

14476 Potsdam-Golm, Germany

Tel: (49)331 567 8112

Fax: (49)331 567 8134

Email: msteinhauser@mpimp-golm.mpg.de

Supporting Tables

Supplemental Table S1. Overview of the optimized enzyme assays, their EC number, the abbreviations used in this work and the pathway they belong to.

Enzyme	EC number	Abbreviation	Pathway
Shikimate dehydrogenase	1.1.1.25	ShkDH	Amino acid metabolism (AAM)
Aspartate aminotransferase	2.6.1.1	AspAT	
Alanine aminotransferase	2.6.1.2	AlaAT	
NAD-dependent glutamate dehydrogenase	1.4.1.3	NAD-GlDH	
NAD-glyceraldehyde 3-P dehydrogenase	1.2.1.12	NAD-GAPDH	Glycolysis / gluconeogenesis (GGP)
Triose phosphate isomerase	5.3.1.1	TPI	
Glucose 6-P 1-dehydrogenase	1.1.1.49	G6PDH	
Phosphoglycerate kinase	2.7.2.3	PGK	
Phosphoenol pyruvate carboxylase	4.1.1.31	PEPC	
Pyrophosphate: fructose 6-P 1-phosphotransferase	2.7.1.90	PPI-PFK	
NADP-glyceraldehyde 3-P dehydrogenase	1.2.1.13	NADP-GAPDH	
Fructose-1,6-bisphosphate aldolase	4.1.2.13	Aldolase	
Pyruvate kinase	2.7.1.40	PK	
ATP-phosphofructokinase	2.7.1.11	ATP-PFK	
Glucokinase	2.7.1.2	GlcK	Branch between glycolysis/gluconeogenesis and sucrose and starch metabolism GGP / SSM
Phosphoglucomutase	5.4.2.2	PGM	
Phosphoglucose isomerase	5.3.1.9	PGI	
Sucrose synthase	2.4.1.13	SuSy	
UDP-glucose pyrophosphorylase	2.7.7.9	UGP	Sucrose and starch metabolism (SSM)
Acid invertase	3.2.1.26	Invertase	
Fructokinase	2.7.1.4	FruK	
Sucrose phosphate synthase	2.4.1.14	SPS	
ADP-glucose pyrophosphorylase	2.7.7,27	AGP	
Aconitase	4.2.1.3	Aconitase	Tricarboxylic acid cycle (TCA)
NAD-malate dehydrogenase	1.1.1.37	NAD-MDH	
Fumarate hydratase	4.2.1.2	Fumarase	
NADP-Isocitrate dehydrogenase	1.1.1.42	NADP-IcDH	
Succinyl CoA ligase	6.2.1.5	SCS	

Supplemental Table S2. Trait heritability of metabolite traits in the *S lycopersicum* `M82´ x *S. pennellii* introgression line population.

	IL 2003			IL 2004			Heritability H^2			Correlation	
Metabolites	**ØCV [within]**	**CV [among]**	H^2	**ØCV [within]**	**CV [among]**	H^2	**Ø**	**CV**	**CL**	r	r_s
amino acids	25	23	31	38	21	10	20.4	75	↔	0.5**	0.49**
glucose	19	15	23	18	14	28	25.5	11	↔	0.53**	0.47**
sucrose	45	40	19	34	14	0	9.6	141	↓	0.01	-0.07
fructose	24	14	5	12	6	6	5.7	8	↓	-0.07	-0.07
proteins	22	19	28	31	19	13	20.5	51	↔	0.41**	0.44**
Average	27	22	21	27	15	11	16.3	57	↓	0.28	0.25

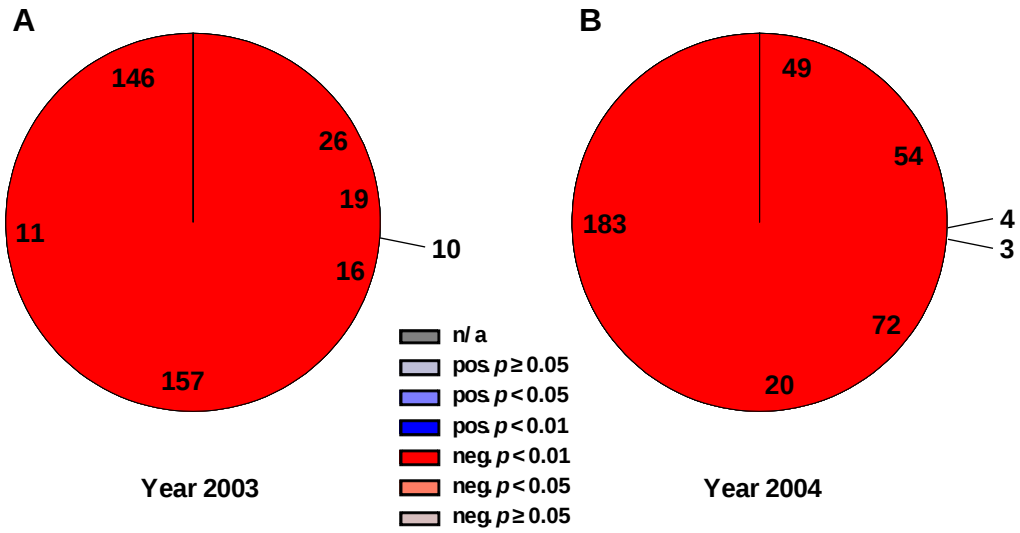
The coefficient of variation (CV) in percentage within and among the lines as well as the heritability (H^2) for each enzyme activity trait is presented for the two independent field trials 2003 and 2004. For each trait the mean-average CV (ØCV) exhibits the average of CV values obtained for each introgression line (IL); the CV among lines was computed using the mean enzyme activities among the lines. The columns Ø and CV show the average heritability and the corresponding CV, respectively. The class (CL) exhibits a grouping of the average H^2 values with: ↓ = low ($H^2 \leq 20$), ↔ = intermediate ($20 < H^2 \leq 40$), and ↑ = high ($H^2 > 40$) heritability. Pearson's (r) and Spearman's (r_s) correlation and its significance (* - $P < 0.05$; ** - $P < 0.01$) of enzyme activity levels in the ILs between the two field trials are also displayed.

Supplemental Table S3. Trait heritability of selected metabolite traits in the *S lycopersicum* `M82´ x *S. pennellii* introgression line population, analyzed using GC/MS-based metabolite profiling by Schauer et al. (2008).

			Heritability H^2					Correlation
Metabolites	**Substrat of**	**Product of**	**IL 2003**	**IL 2004**	**Ø**	**CV**	**CL**	***r***
citrate	Aconitase		22	40	31	41	↔	0.43
fructose	SuSy; FruK	Invertase	40	48	44	13	↑	0.18
glucose	GlcK	Invertase	43	40	42	5	↑	0.23
aspartate	AspAT		35	21	28	35	↔	0.52
sucrose	Invertase	SuSy	21	28	25	20	↔	0.47
malate		Fumarate; NAD-MDH	11	22	17	47	↓	0.45
glutamate	GlDH		22	13	18	36	↓	0.23
succinate		SCS	36	21	29	37	↔	0.28
isocitrate	NADP-IcDH; Aconitase		21	38	30	41	↔	0.30
3-phosphoglyceric acid	PGK		54	28	41	45	↑	0.18
fructose 6-P	PPi-PFK; SPS; PGI; ATP-PFK	FruK	34	19	27	40	↔	0.27
alanine	AlaAT		30	19	25	32	↔	0.08
glucose 6-P	G6PDH	GlcK; PGI; PGM	28	21	25	20	↔	0.27
aconitate		Aconitase	19	n/d	19	n/d	↓	n/d
2-oxoglutarate		GlDH	47	n/d	47	n/d	↑	n/d
shikimate	ShkDH		12	12	12	0	↓	-0.13
Average			30	26	28	29	↔	0.27

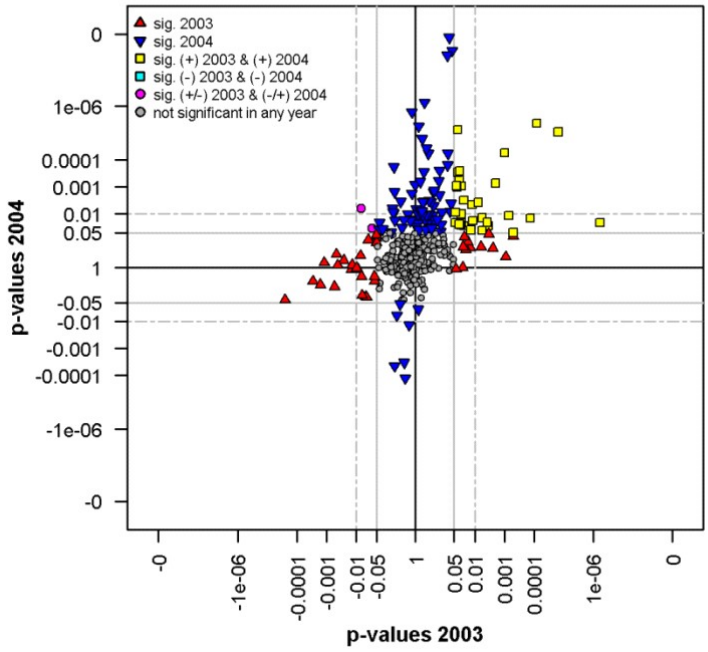
The heritability (H^2) for each metabolite trait is presented for the two independent field trials 2003 and 2004. The columns Ø and CV show the average heritability and the corresponding CV, respectively. The class (CL) exhibits a grouping of the average H^2 values: ↓ = low ($H^2 \leq 20$) ↔ = intermediate ($20 < H^2 \leq 40$), and ↑ = high ($H^2 > 40$) heritability. Pearson's (r) correlations of metabolites levels in the ILs between the two field trials are also displayed. All data were taken from Schauer et al. (2008) and metabolite traits selected according to their substrate or product relationship with respect to enzyme activities analyzed in this study.

Supporting Figures

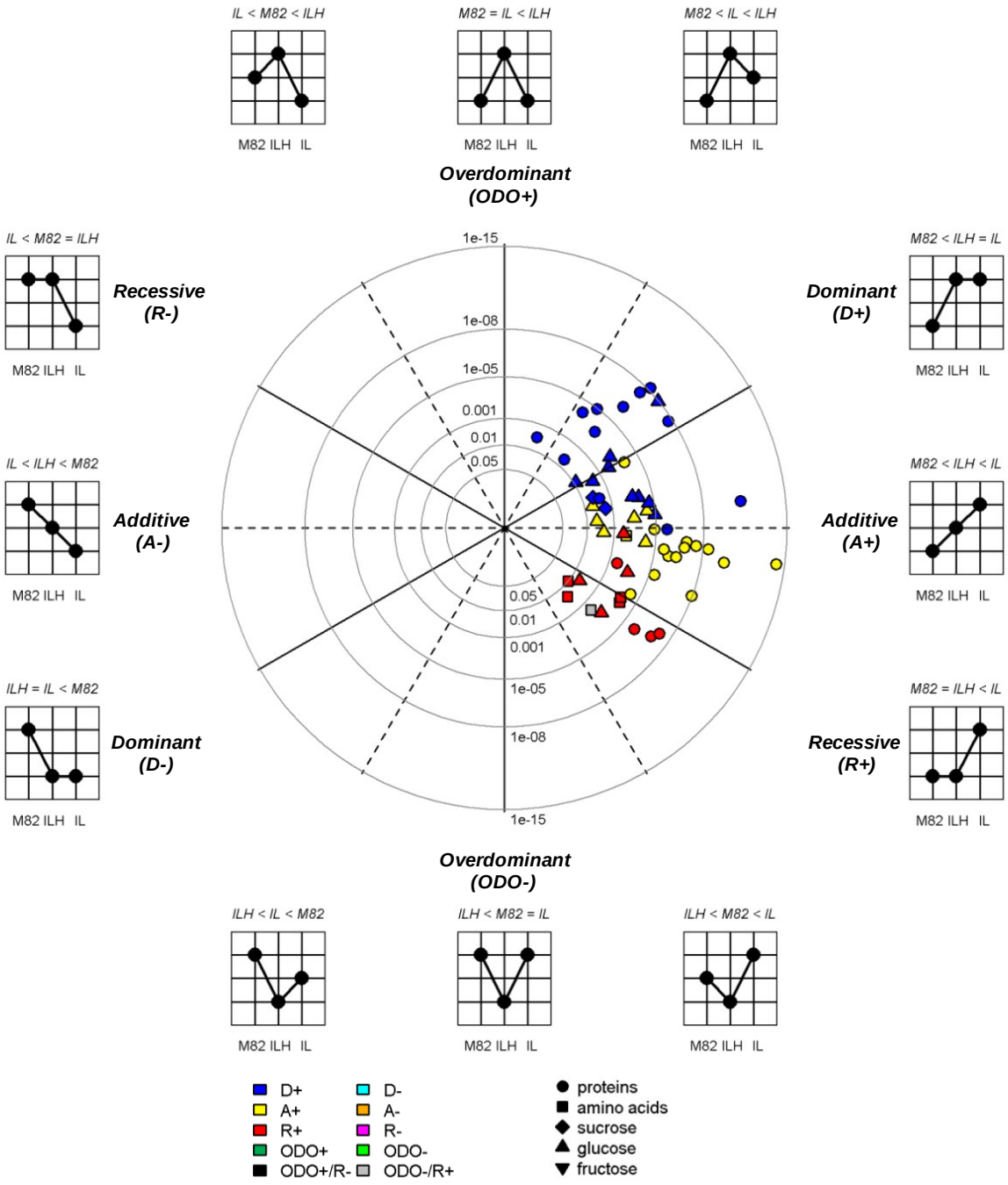


Supplemental Figure S1. Distribution of *P*-values derived from *t*-test analyses of metabolites observed in the homozygote introgression lines for the field trials (A) 2003 and (B) 2004. Blue colored sectors represent the number of positive traits, i.e. the introgression revealed higher values than the parental control `M82´. Red colored pie sectors reflect negative traits where the observed value in the introgression is lower than the parental control. Grey sectors depict the portion of *t*-tests that were not conducted as less than three replicates were available with respect to genotype and enzyme activity. The observed *P*-values are grouped accordingly as depicted in the figure with: (i) dark red / dark blue color - significant portion in range of $0 \leq P < 0.01$, (ii) red / blue - significant portion in range of $0.01 \leq P < 0.05$, and faded light red / faded light blue - not significant portion of $P \geq 0.05$.

5



Supplemental Figure S2. Scatter plot of the *P*-value distribution derived from *t*-test analyses of metabolites observed in the homozygote introgression lines for the field trials 2003 and 2004. The *P*-values were computed separately by *t*-tests for the data set form each year. Traits were considered significant at $P < 0.05$. Observed positive or negative traits in an IL line compared to the parental control `M82´ are reflected by the sign of the *P*-values. To aid visualization *P*-values were inverted and $\log_{10}$-transformed to separate significant from non-significant effects. The significance levels of $P < 0.05$ and $P < 0.01$ are depicted as solid and dotted lines, respectively. Traits are represented by colored shapes as depicted, with: (i) yellow-colored squares – positive significant traits ($P < 0.05$) in both independent experiments, (ii) cyan-colored squares – negative significant traits ($P < 0.05$) in both experiments, (iii) magenta-colored circles – positive significant in one trial and negative significant in the other experiment, (iv) blue-colored lower triangles – significant traits only in the experiment 2004, (v) red-colored upper triangle significant traits only in the experiment 2003, (vi) grey-colored circles – no significant traits in any of the two independent experiment. Data with less than three replicates per genotype and metabolite in any of the trials and ILs analyzed only in one experiment were excluded.

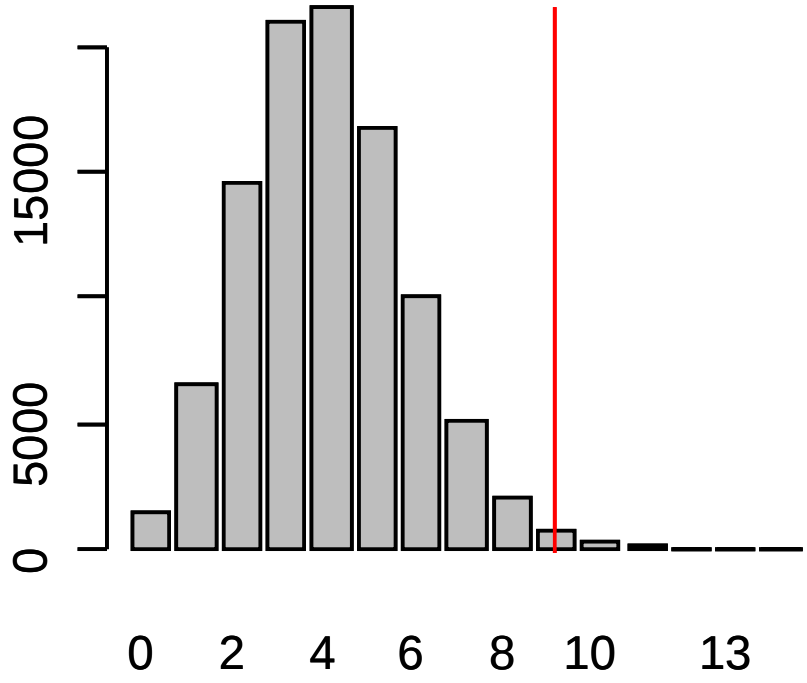


87

88**Supplemental Figure S3. Two-dimensional polar plot representation of the mode of**
89**inheritance and associated *P*-values of detected metabolite QTLs estimated using**
90**the phenotypic effects in the homozygote (ILs) and heterozygote (ILHs)**
91**introgression lines and the parental control `M82´.** The data are from the field trial of
922004. Traits positioned on the dashed black lines exhibit metabolite mean differences of
93one genotype which is exactly in the middle between the genotypes with low and high
94phenotypic effects. Traits exhibiting clear additive (A) or over-dominant (ODO) effects are

located on the horizontal and vertical lines, respectively. The distance to the center (radius) reflects the *P*-value associated with a trait estimated using ANOVA of the corresponding homozygote IL data measured for both, the 2003 and 2004 trials. The shape of the plotted traits corresponds to the metabolite as depicted in the figure legend. The color of each shape (see figure legend) corresponds to the mode of inheritance of the trait classified using a decision tree suggested by Semel et al. (2006). Only traits are visualized which were detected and evaluated using both, *t*-test analyses on the IL / ILH data from 2004 and ANOVA on IL data from 2003 and 2004 (see Material and Methods, Supplemental Data S3).

8



Supplemental Figure S4. Estimation of the frequency of spurious collocations of enzyme activity QTL and structural genes for that enzyme. To asses the number of co-locations that might occur by chance, a strategy similar to that of Lisec et al. (2008) was adopted. The locations of all annotated genes for the considered enzymes were determined based on the tomato genome release (see Material and Methods). If multiple isoforms (e.g. tandem duplications) were located on the same IL, they were counted as one enzyme for the sake of the simulation, as a separation would require a better resolution for the QTL locations, giving 27 QTL. 27 QTL for the 13 enzymes were randomly sampled 100,000 times from a population of 72 ILs, and the number of co-locations with the observed QTL was counted and represented in histogram form. The number of observed co-locations (9) is indicated by the red bar. The median of the distribution of chance co-locations is 4 co-locations, and the 95% quantile is 7. A similar conclusion was reached when the frequency distribution was generated by randoming QTL rather than gene location (not shown) or when an underlying binomial distribution was assumed (not shown).

9

Supporting Data

Supplemental Data S1: Overview of processed samples per genotype and considered mean-average values per genotype and assay.

The sheet '*Genotypes*' contains *per*-genotype information with respect to the number of processed samples per genotype and field experiment ('*Samples per genotype*'). The samples used for the reference genotype `M82´ are exactly the same for the homozygote (IL) and heterozygote (ILH) introgression lines in the year 2004. The number ('*Count*') and frequency ('*Frequency [%]*' in percentage) of considered mean-average values (n ≥ 3 replicates per combination of genotype and assay) are depicted per genotype and field experiment ('*Assays with* $n_{Replicates} \geq 3$'). In total, 28 enzyme activity assays (cf. Supplemental Table S1) and 5 metabolite assays (total content of amino acids, proteins, fructose, glucose and sucrose) were performed per sample.

The sheet '*Assays*' contains the *per*-assay information with the number ('*Count*') and frequency ('*Frequency [%]*') of considered mean-average values per assay and field experiment. Additionally, the number of genotypes processed per field experiment is depicted ('*Lines (incl. `M82´)*').

cf. *Steinhauser-TomQTL-Supplemental-Data-S1.xls* (attached MS Excel file)

Supplemental Data S2: Mean-average values and *t*-test statistics of maximal enzyme activities and metabolite pool sizes, determined on tomato fruit pericarp tissue harvested at ripe stage of fruit development, of *S. pennellii* introgression lines compared to the reference genotype *S. lycopersicum* `M82´.

The sheet '*Enzymes*' contains the information derived from enzyme activity assays, the sheet '*Metabolites*' the information related to metabolite assays. To aid interpretation all lines are depicted independent of whether sample material was available within a respective year. A '*n/a*' cell content indicates that the genotype was not processed in a particular year; blank content indicates that computations were not conducted due to a limited number of replicated measurements, i.e. $n < 3$

For each genotype and enzyme / metabolite the mean-average ('*Mean*'), standard deviation ('*SD*') and number of considered (i.e. after outlier-removal) values ('*N*') are depicted per field experiment and split according homo- / heterozygote introgression lines (IL and ILH, respectively). The $\log_2$-transformed ratio ('log_2R') and the uncorrected *P*-values ('*P(IL)*' or '*P(ILH)*') derived from *t*-test statistics per IL in comparison to the reference genotype `M82´are also provided.

cf. *Steinhauser-TomQTL-Supplemental-Data-S2.xls* (attached MS Excel file)

Supplemental Data S3: Overview of QTL mapping and mode of inheritance statistics for maximal enzyme activities and metabolite pool sizes, determined on tomato fruit pericarp tissue harvested at ripe stage of fruit development, for the *S. lycopersicum* x *S. pennellii* introgression line population.

The sheet '*Enzymes*' contains the information derived from enzyme activity assays, the sheet '*Metabolites*' the information related to metabolite assays. All information is depicted on a genotype and enzyme / metabolite basis. To aid interpretation all lines are depicted, independent of whether material was available within a respective year. A '*n/a*' cell content indicates that the genotype was not processed in a particular year; blank content indicates that computations were not conducted due to a limited number of replicated measurements, i.e. $n < 3$.

The table '*ANOVA: IL 2003/2004*' in each sheet contains the *P*-values ('*P*') and the percentage variation ('*%*') for the genotype ('*(G)*'), environment ('*(E)*') and genotype x environment interaction ('*(GxE)*') effect for the homozygote ILs of the field experiment 2003 and 2004. Also, the considered QTLs ('*QTL (P<.01)*' i.e. $P(G) < 0.01$ and $P(GxE) \geq 0.01$ are provided, with comments ('*QTL info*') if the observed changes within both years are in different direction.

The table '*t-test: IL 2003 and IL 2004*' in each sheet contains the *response* ('*R(IL)*'), expressed as percentage difference to the reference genotype `M82´ with the sign indicating the direction, and the *P*-value ('*P(IL)*') derived from *t*-test statistics. Also, the considered QTLs ('*QTL (P<.05)*'), i.e. $P(IL) < 0.05$ in both years with effects in same direction, are depicted within the comment column ('*QTL info*'). For further details see Supplemental Data S2.

The table '*Inheritance (IL / ILH 2004)*' in each sheet contains the *response* of the homozygote ('*R(IL)*') and heterozygote ('*R(ILH)*') introgression line expressed as percentage difference to the reference genotype `M82´, with the sign indicating the

direction. *P*-value derived from *t*-test statistics are depicted for all comparisons, i.e. IL vs. `M82´ (‘*P(IL/P)*’), ILH vs. `M82´ (‘*P(ILH/P)*’), and IL vs. ILH (‘*P(IL/H)*’). The mode of inheritance estimated using decision tree is provided in the column ’*mode (P<.05)*’ at significance level of $P < 0.05$ with: A+ / A- = positive / negative additive, R+ / R- = positive / negative recessive, D+ / D- = positive / negative dominant, and ODO+ / ODO- = positive / negative over-dominant. ODO-/R+ or ODO+/R- depict lines harboring two QTLs with the respective mode. Only the mode of inheritance for those QTLs was considered in further analyses which are supported by ANOVA (‘*ANOVA: IL 2003/2004*’) or *t*-test (‘*t-test: IL 2003 and IL 2004*’) among the ILs of both field experiments, i.e. tagged as enzyme activity (‘*aQTL*’) or metabolite (‘*mQTL*’) QTL, summarized in column ‘*QTL*’.

cf. *Steinhauser-TomQTL-Supplemental-Data-S3.xls* (attached MS Excel file)

Supplemental Data S4: Co-localization analyses of enzyme activity, metabolite, and expression QTLs with structural genes.

Results are depicted for the 27 enzyme activity QTLs (aQTL) identified in this study sorted according to the genotype, i.e. introgression line. For each aQTL the colocation of metabolite (mQTLs; Schauer et al., 2006) and expression QTLs (eQTLs; Baxter et al., 2006) are depicted. If the fields are left blank no information is available. Only mQTLs for substrates or product of all enzymes analyzed in this study are depicted; eQTLs are shown only for significant change in the expression of transcripts encoding the corresponding enzyme. Additionally the co-localization of eQTLs with structural genes are show based on the analysis by Causse et al. (2004) and a refined mapping conducted in this study (cf. Materials and Methods). If a structural gene for a respective enzyme is located within the introgressed *S. pennelii* region the table cell is set to ‘yes’ otherwise ‘no’; sequences in close proximity are marked yellow but set to ‘no’. Furthermore, comments regarding the mapping are provided.

cf. *Steinhauser-TomQTL-Supplemental-Data-S4.xls* (attached MS Excel file)

14

Supplemental Data S5. Analysis of polymorphisms in structural genes that potentially co-locate with enzyme activity QTL and of small gene families where no co-located QTL was found. For each enzyme co-located to its respective QTL, the respective protein and nucleotide sequence were extracted from the `Heinz´ genome and compared to an early draft of the *S. pennellii* genome sequence (Fernie, Usadel, Carrari et al., unpubl.). SNPs causing amino acid changes are shown and potential indels and frameshifts are indicated. To provide an assessment of the importance of these changes, prosite patterns for the respective enzymes are included, as these often cover the active site. To provide a background assessment, the same analysis was performed for small gene families of enzymes that are not co-located. Weak candidates for an enzyme function (e.g. isoforms potentially involved in converting other metabolites) are marked by a “?”.

cf. *Steinhauser-TomQTL-Supplemental-Data-S5.xls* (attached MS Excel file)

15