Current Biology, Volume *23*

Supplemental Information

Evolution of *mir-92a* Underlies Natural Morphological Variation in *Drosophila melanogaster*

Saad Arif, Sophie Murat, Isabel Almudi, Maria D.S. Nunes, Diane Bortolamiol-Becet, Naomi S. McGregor, James M.S. Currie, Harri Hughes, Matthew Ronshaugen, Élio Sucena, Eric C. Lai, Christian Schlötterer, and Alistair P. McGregor

Supplemental Inventory

Figure S1. Raw naked valley areas are plotted against femur length for three of the five populations included in Figure 1C. Mapping strains and individuals from ten different *D. simulans* populations from Madagascar are also included.

Figure S2. This figure includes the QTL map for naked valley variation (A), the naked valley sizes of recombinant flies that excluded the involvement of *Ubx* (B) and intermediate resolution map (C) ultimately leading to the 25kb resolution map shown in Figure 2**.**

Figure S3. This figure is linked to Figure 3 in the main text. This figure shows, along with Figure 3, that the ectopic expression of *mir-92a* induces equivalent phenotypic effects in wings depending on the GAL4 driver used.

Figure S4. This figure provides a pictorial representation of sequence conservation in and around the *mir-92a* sequence in our focal strains (A) and to support our reasoning that causative changes are of a regulatory nature, evidence for which is provided in Figure 4. (B) is a pictorial representation of the sequence conservation in miR-92a target sites in the *sha* 3'UTR across *Drosophila* species. This evolutionary conservation highlights the possible functional significance of the regulatory relationship between *sha* 3'UTR and miR-92a which is evidenced in Figure 4 in the main text.

Table S1. This table provides a summary of the results of the QTL analyses that are described in the main text.

Table S2. This table provides information on the molecular markers used in this study and feature in the results of Figure 2 and Figure S2A and C.

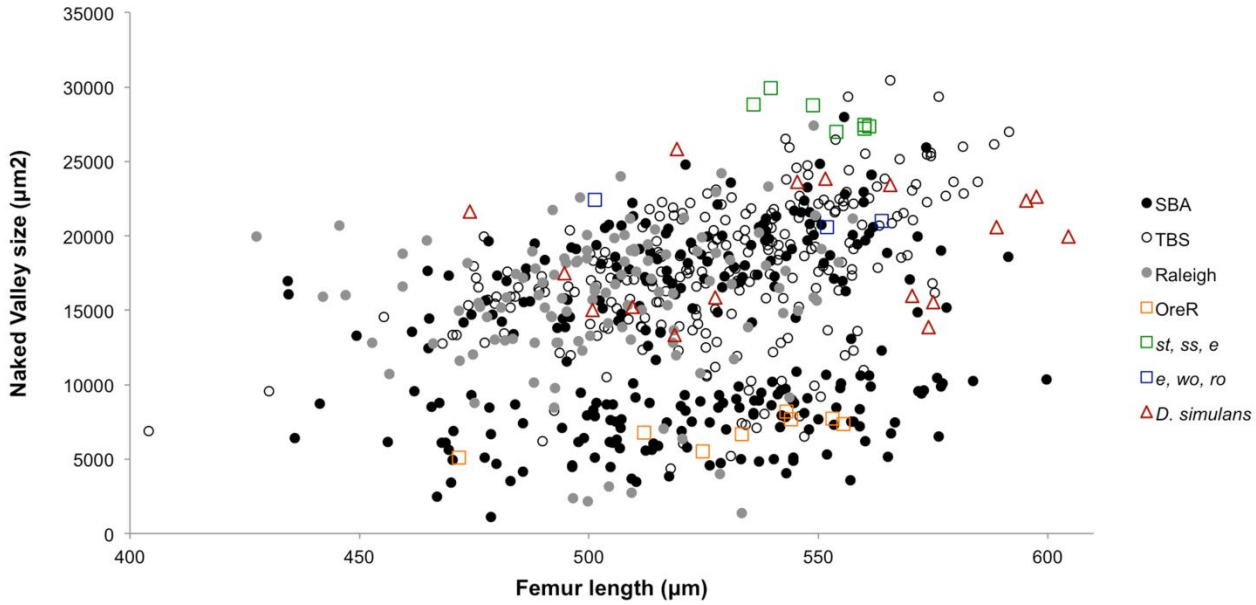

Figure S1. Distribution of Naked Valley Sizes among *D. melanogaster* Populations (Related to Figure 1)

Scatter plot of naked valley area against femur length of individuals from *D. melanogaster* populations collected in Spain (SBA), Turkey (TBS) and North America (Raleigh). Strains used for mapping purposes (OreR, *e, wo, ro and st, ss, e*), as well as *D. simulans* individuals, are shown for comparison.

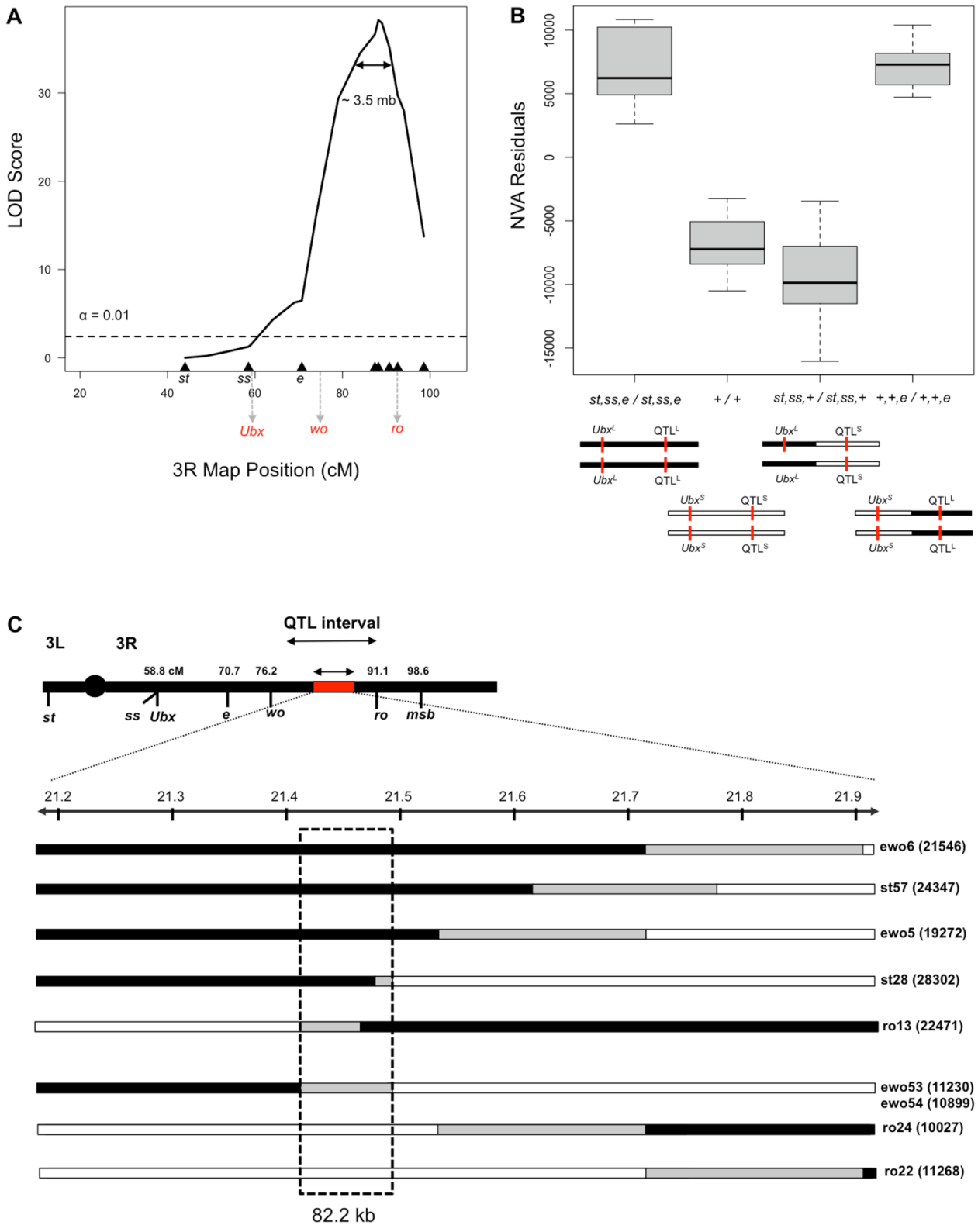
A
LOD Score
~ 3.5 mb
α = 0.01
st
ss
e
Ubx
wo
ro
3R Map Position (cM)
B
NVA Residuals
st,ss,e / st,ss,e
+ / +
st,ss,+ / st,ss,+
+,+,e / +,+,e
C
QTL interval
3L
3R
58.8 cM
70.7
76.2
91.1
98.6
st
ss
Ubx
e
wo
ro
msb
ewo6 (21546)
st57 (24347)
ewo5 (19272)
st28 (28302)
ro13 (22471)
ewo53 (11230)
ewo54 (10899)
ro24 (10027)
ro22 (11268)
82.2 kb

Figure S2.

Figure S2. Mapping the Evolved Region Responsible for Variation in Naked Valley Size (Related to Figure 2)

(A) Identification of a large effect QTL for naked valley variation on chromosome 3R. LOD profile of a QTL analysis of backcross progeny from a *st, ss, e/*Oregon R F1 to *st, ss, e* cross. Femur length was used as a covariate in this analysis. There is a single peak at 88.2 cM with the double-headed arrow indicating the 2-LOD support interval for the QTL. The dashed line represents the significance threshold based on 1000 permutations. Marker positions are represented by black triangles on the x-axis. Map positions of *Ubx* and markers used for fine scale mapping (*wo*, *ro*) are highlighted by the dashed gray arrows.

(B) Variation in naked valley area in *D. melanogaster* is not caused by *Ubx*. Boxplots display the naked valley phenotype (residuals of naked valley area regressed on femur length) for two parental genotypes (*st, ss, e* and Oregon R) and two reciprocal homozygous recombinant lines generated from crosses between parental strains to generate recombination events between *Ubx* and *e*. 7 to 10 flies were measured for each genotype. A linear model (ANOVA Type II SS, data not shown) fit to this data set shows no significant contribution (*p-value = 0.1998*) of the proximal region of 3R (containing *Ubx*) while the distal region (containing our QTL) has a significant effect ($p < 0.0001$) and accounts for ~87% of the phenotypic variation in this data set consistent with the effect of the QTL detected in this region.

(C) Mapping of the evolved region to a resolution of 82 kb. The upper black bar represents chromosome 3 with the two arms (3L and 3R) indicated either side of the centromere (circle). The QTL region in (A) is indicated by a double-headed arrow. The position of *Ubx* and selected QTL markers are shown below the bar with their positions in centimorgans indicated above (msb = microsatellite b). The red bar represents the 0.7 Mb region shown expanded between the broken diagonal lines with the scale showing the Mb positions on 3R. The bars below the scale indicate selected recombinants with breakpoints in the 0.7 Mb region (note that all flies also carried a non-recombinant chromosome either from strain *e, wo, ro* or *st, ss, e* that is not shown). Recombinant breakpoint positions determined by molecular markers are indicated by black triangles. The identity of each recombinant fly and its naked valley size in μm^2 is shown on the right. Chromosomal regions from strains *e, wo, ro* or *st, ss, e* (large naked valley parental lines used) and Oregon R (small naked valley parental line) are indicated in black and white respectively. DNA regions indicated in grey represent regions where the parental strain identity was not determined. The box indicates the 82.2 kb resolution of the evolved region determined by these recombinant flies.

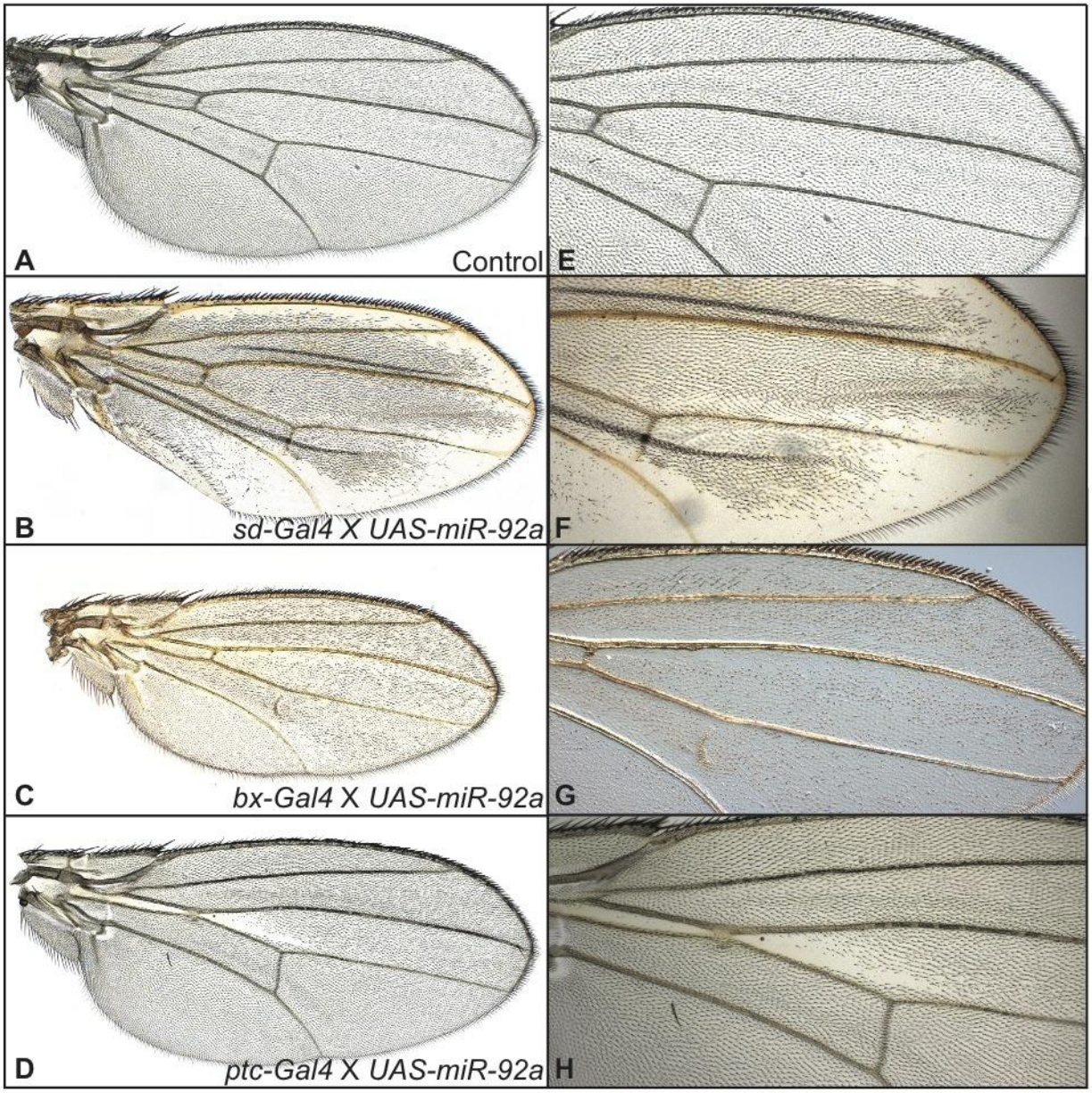

Figure S3. Ectopic Expression of *mir-92a* Represses Wing Trichomes (Related to Figure 3)

(A-D) Morphology of control wing (A) and effects of *mir-92a* over expression in developing wings driven by *sd-Gal4* (B), *bx-Gal4* (C), and *ptc-Gal4* (D).

(E-H) Higher magnification images of wings in (A-D) showing loss of wing trichomes.

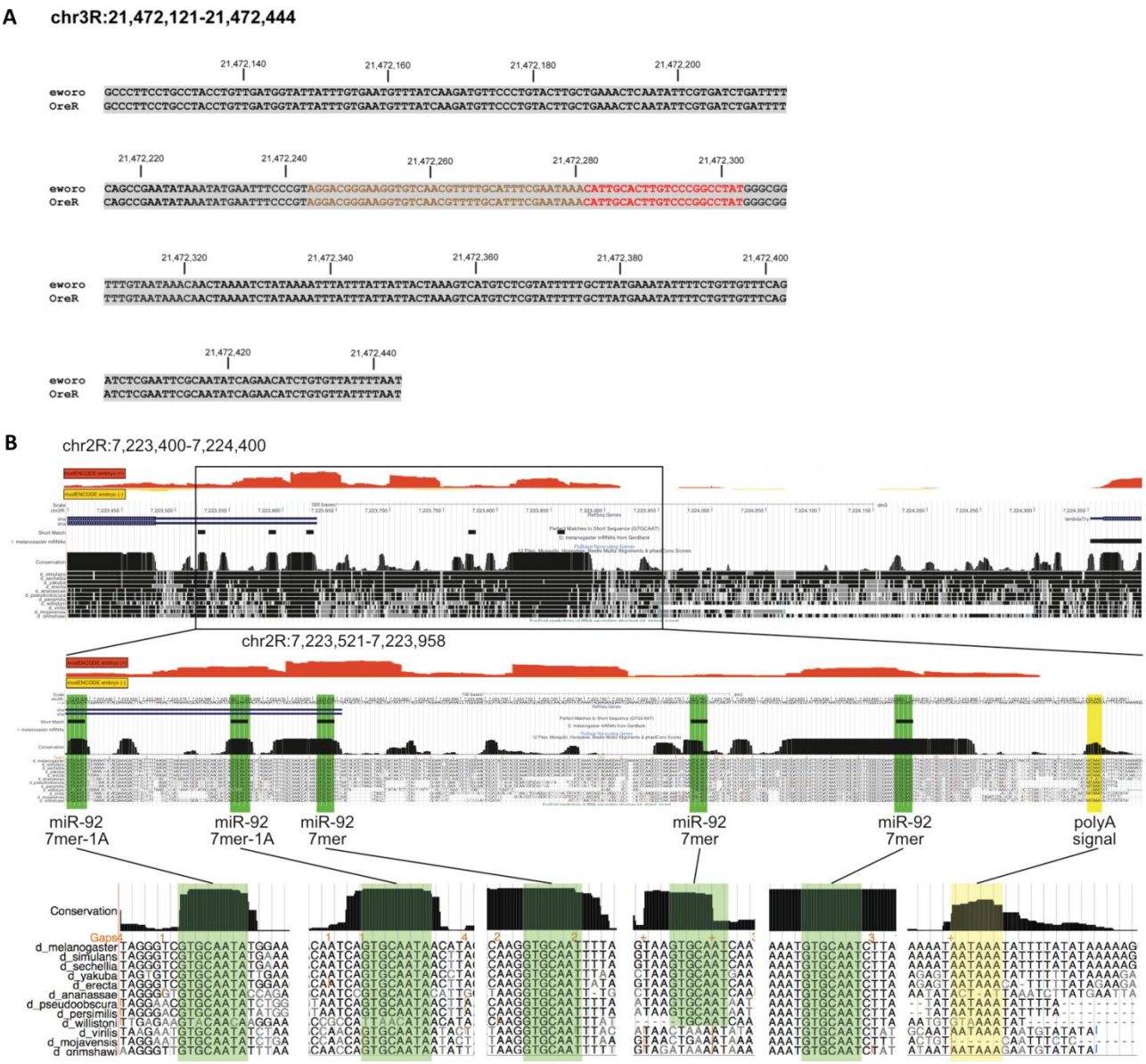

Figure S4. Sequence Conservation of miR-92a Sequence and Target Sites in the *sha* 3'UTR across *Drosophila* Genomes (Related to Figure 4)

(A) The 22 nt sequence and flanking regions of *mir-92a* are conserved between *D. melanogaster* strains with large and small naked valleys. Alignment of the *mir-92a* locus from *e, wo, ro* and Oregon R strains. The mature *mir-92a* sequence is highlighted in red. The pre-miRNA sequence is highlighted in brown.

(B) miR-92a target sites are highly conserved in *sha* 3'UTR. UCSC sequence alignment shows that the target sites for miR-92a in *sha* 3'UTR (outlined in green) are highly conserved among 12 *Drosophila* species.

Table S1. QTL Analysis and Effect Size Summary of Two Different Backcross Progeny Populations Scored for Naked Valley Area

Cross	Sample Size	QTL Position (cM)[1]	LOD Score[1]	2-LOD CI (cM)[1]	% Variation QTL[1, 2]	% Variation FL[3]	Additive Effect ± SE (μm^2)[4]	Relative Homozygous Effect[5]
st,ss,e x OreR	180	88.2 (88.2)	38.2 (29.2)	84–90.7 (84–92.6)	57.1 (50.8)	12.5	-9760 ± 718	-0.91
e,wo,ro x RAL514	195	85.7 (88.7)	32.8 (20.6)	79.7–89.7 (81.7–91.1)	41.5 (39.4)	24.7	-4953 ± 330	-0.69

[1]Numbers outside parenthesis correspond to results from an analysis with femur length as a covariate, whereas numbers in parenthesis are from a model without accounting for any femur size differences.

[2]Percentage of phenotypic variation in naked valley area accounted for by the QTL in the backcross population.

[3]Percentage of phenotypic variation in naked valley area accounted for by femur length (FL) in the backcross population.

[4]Estimated as the difference between naked valley areas between homozygotes for the large QTL allele and heterozygotes with a single large and single small QTL allele.

[5]The additive effect standardized by half the difference between parental averages.

Table S2. Markers Used for Genotyping

Marker Name	Position (Mb)	Diagnostic	Forward Primer	Reverse Primer
f	21231144-21231179	microsatellite	TCCATGCGGTAATGAAATCC	ACGCATCGTTGTTTGCACAT
889	21410889-21411482	RFLP (SpeI)	AATGAGCGGCATAAACGAAT	TTTCGGAGGAGAGCACTAGC
6543	21426543-21427085	SNPs	CACCCACAACCACTACCACA	GCGCCATTCTTGGTTATACG
7652	21436847-21437570	SNPs	ACGAACCGCCTCTAAGTGTG	GCTGCAACCCATTTACACCT
CG5071_2	21447189-21447898	SNPs	ATTGTGATTGGCTTGGCAGT	TCGACATCCAAATCCGTGTA
569	21451550-21452298	SNPs	CCAGGGCAGTAAGAGGAGTG	GAGGATTCTCGGCACTGGT
mir-92a_HpaI	21472408-21473187	RFLP (Hpa I)	CGAATTCGCAATATCAGAACA	ATTAAGATTGCCTGCGCAAC
mir-92b_PstI	21476847-21477561	RFLP (Pst I)	TCCTTTTTCTTTAATAATGCGTCT	ATTACAGGGCCAGACATTCG
896	21490788-21491587	RFLP (HinfI)	CCAATCTCTACAATTCGCCATT	GCGAAGGATATCGACAAGGA
3516	21483516-21484093	SNPs	ATTGAGTGGCGTGAGTAGGG	TTGAGATTTCCTTGGGCAAC
3137	21493137-21493707	SNPs	TTAAAGGGTTCCTTCGCTTG	CACAAATCCTGGCTCTTTCC
f+0.3	21531917-21532494	SNPs	ATGGCACCGCTTTGTTGT	CGCGTAAGTGTTGCTTGAAA
2161	21616592-21617435	SNPs	TTTCCTGCACTTTTCCCTTC	CTCTCGGGACTTGACAGCAT
g-1.71	21704404-21704904	SNPs	TCGATGCGGTTAATTGTGG	TCTTGAATTTGTTCATGTGTTGA
g-1.1	21774297-21774788	SNPs	TAAAAGACGCTCGGAGTTGC	CTGGAAAACTGTCGCTTCGT
g	21875074-21875097	microsatellite	ACAGCGAATGCAACAACAAA	CAATGCGAGTGAGTGTTTCG
g+0.3	21878373-21879038	SNPs	TTACAATGTGTGTGTGTTGTATTGT	CATGTTCTGCATGGCCTTC
d	22523838-22523854	microsatellite	GCCGTCACCTCTTCATGTG	AACGGTGAGCAGAGAGATACAGA
c	23528849-23529019	microsatellite	TGGGTCTGCGACTCAAA	GTTAGCCGGCCCTAGATACT
b	25189492-25189516	microsatellite	GGCCAATGTGGGTGAGGT	CATCTCTGCGCTTGATCCTT