Mating resource assessment behaviour exists in many species[17–20]. Bluehead wrasses may not exercise such assessment because there may be few or no fitness differences among potential sites[21]. In this case, the time saved and safety gained through traditional behaviour may well outweigh any benefits derived from frequent careful assessment[22]. However, traditions can continue in the same area after a resource base is altered because they are, by definition, less responsive to external conditions[5–7]. This study should inject a note of caution into attempts to test hypotheses about the evolution of mating systems by measuring the current spatial distribution of resources or females[23–25].

I thank T. Boundy, T. Brown, K. Clifton, L. Clifton, D. Wiley, J. Crisp, L. Fore, P. Hart, J. Jolly, J. and V. Maldonado, G. McKinley, B. McAffee, S. Olivier, M. Schildhauer and E. Schultz for field assistance. K. Clifton, E. van den Berghe, J. Endler, S. Rothstein and E. Schultz read the manuscript. The Kuna people of the Comarca de San Blas, Panama, kindly provided permission to study in their water. This work was supported by the NSF.

Received 16 May; accepted 1 September 1988.

1. Parker, G. A. in *Mate Choice* (ed. Bateson, P.) 141–166 (Cambridge University Press, Cambridge, 1983).
2. Partridge, L. & Halliday, T. R. in *Behavioural Ecology, an Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 222–250 (Sinauer, Sunderland, 1984).
3. Emlen, S. T. & Oring, L. W. *Science* **197**, 215–223 (1977).
4. Vehrencamp, S. L. & Bradbury, J. W. in *Behavioural Ecology, an Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 251–278 (Sinauer, Sunderland, 1984).
5. Parker, G. A. *A. Rev. Ent.* **23**, 173–196 (1978).
6. Bonner, J. T. *The Evolution of Culture in Animals* (Princeton University Press, 1980).
7. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process* (University of Chicago Press, 1985).
8. Galef, B. G. Jr. *Adv. Study Behav.* **6**, 77–100 (1976).
9. Mainardi, D. in *Sociobiology: Beyond Nature/Nurture?* (eds Barlow, G. W. & Silverberg, J.) 227–255 (Westview, Boulder, 1980).
10. Warner, R. R. & Robertson, D. R. *Smithson. Contr. Zool.* **254**, 1–27 (1978).
11. Robertson, D. R. *Copeia* **1987**, 637–642 (1987).
12. Warner, R. R. *Anim. Behav.* **35**, 1470–1478 (1987).
13. Victor, B. C. *Ecol. Monogr.* **56**, 145–160 (1986).
14. Randall, J. E. & Randall, H. A. *Zoologica (N.Y.)* **48**, 49–60 (1963).
15. Helfman, G. S. & Schultz, E. T. *Anim. Behav.* **32**, 379–384 (1984).
16. Warner, R. R. *Evolution* **38**, 148–162 (1984).
17. Searcy, W. A. *Am. Nat.* **114**, 77–100 (1979).
18. Pleszczynska, W. K. & Hansell, R. I. C. *Am. Nat.* **116**, 821–830 (1980).
19. Borgia, G. *Evolution* **36**, 307–315 (1982).
20. Thornhill, R. & Alcock, J. *The Evolution of Insect Mating Systems* (Harvard University Press, 1983).
21. Shapiro, D. Y., Hensley, D. A. & Appeldoorn, R. S. *Envir. biol. Fish.* **22**, 3–14 (1988).
22. Warner, R. R. *Proc. Fifth Int. Coral Reef Congress* **4**, 145–150 (1985).
23. Lill, A. *Z. Tierpsychol.* **36**, 513–530 (1974).
24. Robertson, D. R. & Hoffman, S. G. *Z. Tierpsychol.* **45**, 298–320 (1977).
25. Bradbury, J. W., Gibson, R. & Tsai, I. M. *Anim. Behav.* **34**, 1694–1709 (1986).
26. Warner, R. R. & Hoffmann, S. G. *Evolution* **34**, 508–518.

Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms

Andrew H. Paterson*, Eric S. Lander†‡, John D. Hewitt§, Susan Peterson*, Stephen E. Lincoln† & Steven D. Tanksley*

* Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853, USA
† Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA
‡ Harvard University, Cambridge, Massachusetts 02138, USA
§ Department of Vegetable Crops, University of California, Davis, California 95616, USA

The conflict between the Mendelian theory of particulate inheritance[1] and the observation of continuous variation for most traits in nature was resolved in the early 1900s by the concept that quantitative traits can result from segregation of multiple genes, modified by environmental effects[2–5]. Although pioneering experiments[6–9] showed that linkage could occasionally be detected to such quantitative trait loci (QTLs), accurate and systematic mapping of QTLs has not been possible because the inheritance of an entire genome could not be studied with genetic markers[7]. The use of restriction fragment length polymorphisms[10] (RFLPs) has made such investigations possible, at least in principle. Here, we report the first use of a complete RFLP linkage map to resolve quantitative traits into discrete Mendelian factors, in an interspecific back-cross of tomato. Applying new analytical methods, we mapped at least six QTLs controlling fruit mass, four QTLs for the concentration of soluble solids and five QTLs for fruit *p*H. This approach is broadly applicable to the genetic dissection of quantitative inheritance of physiological, morphological and behavioural traits in any higher plant or animal.

The parents for the back-cross were the domestic tomato *Lycopersicon esculentum* cv. UC82B (denoted E) and a wild South American green-fruited tomato *L. chmielewskii*[11] accession LA1028 (denoted CL). These strains have very different fruit masses (E ~ 65 g; CL ~ 5 g) and concentrations of soluble solids[12] (E ~ 5%; CL ~ 10%)—traits of agricultural importance, because they jointly determine the yield of tomato paste. In addition, the strains are known to be polymorphic for genes affecting fruit *p*H, which is important for the optimal preservation of tomato products[13]; the difference in *p*H between the parental strains is, however, small.

A total of 237 back-cross plants, with E as the recurrent parent, were grown in the field at Davis, California. Between five and 20 fruit from each plant were assayed[13] for fruit mass, soluble-solids concentration (°Brix; see Fig. 1 legend for definition) and *p*H, each of which showed continuous variation (Fig. 1). Soluble-solids concentration correlated negatively with fruit mass ($r = -0.42$) and positively with *p*H ($r = +0.33$).

We had previously constructed a genetic linkage map of tomato[14] with over 300 RFLPs and 20 isozyme markers, by analysing 46 F_2 individuals derived from *L. esculentum* cv. VF36 × *L. pennellii* accession LA716 (E × P). The map is essentially complete: it has linkage groups covering all 12 tomato chromosomes with an average spacing of 5 cM between markers (1 cM is the distance along the chromosome which gives a recombination frequency of one per cent). For QTL mapping, we selected a subset of markers spaced at approximately 20 cM intervals and displaying polymorphism between the E and CL strains. These included 63 RFLPs and five isozyme markers. In addition, the E and CL strains differ in two easily-scored, simply-inherited morphological traits: determinacy (described below) and uniform ripening, controlled by the *sp* and *u* genes, respectively. Although a few distal regions did not contain appropriate markers, we estimate that about 95% of the tomato genome was detectably linked to the markers used.

These 70 genetic markers were scored for each of the 237 E × CL back-cross progeny (as described in ref. 13), and a linkage map was constructed *de novo* using MAPMAKER[15]. The map covers all 12 chromosomes with an average spacing of 14.3 cM. Although the linear order of markers inferred from the E × CL cross essentially agreed with that inferred from the E × P cross (but see Fig. 3 legend), genetic distances differed markedly in certain intervals (for example, 51 cM in E × P and 11 cM in E × CL, for the distance between the 45S ribosomal repeat and *TG1B* on chromosome 2). In total, the markers scored in both crosses span 852 cM in the E × CL map versus 1103 cM in the E × P map, a highly significant ($P < 0.01$) difference. Skewed segregation ($P < 0.05$) was detected for 48 of the 70 markers, comprising 21 distinct regions distributed over all 12 chromosomes. The heterozygote (E/CL) was over-

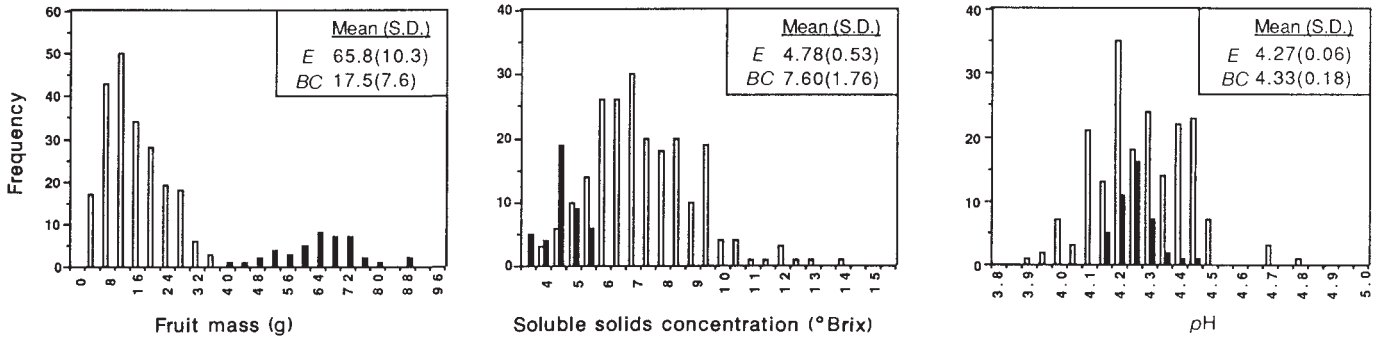
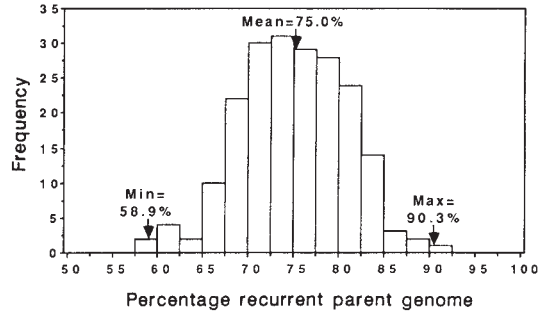


Fig. 1 Frequency distribution for fruit mass, soluble-solids concentration (°Brix, a standard refractometric measure primarily detecting reducing sugars, but also affected by other soluble constituents; 1 °Brix is approximately 1% w/w) and *p*H in the E parental strain and in the back-cross (BC) progeny. The tomatoes were grown in the field at Davis, California, in a completely randomized design including 237 BC plants (with E as the recurrent pistillate parent), as well as E, CL and the F_1 as controls. Neither CL nor the F_1 progeny matured completely, as is typical in the central valley of California. Among the BC plants, six failed to mature and 12 produced too few fruit to assay reliably for quantitative traits. The absence of quantitative trait data for these few progeny should yield at most a slight bias in our analyses. Means and standard deviations for the distributions of the E parental strain (*E* filled bars) and the BC progeny (*BC* open bars) appear in the upper right of each histogram. The distributions for soluble-solids concentration and *p*H are approximately normal. The distribution of the BC progeny for fruit weight is clearly skewed; $\log_{10}$ (fruit mass) was studied throughout to achieve approximate normality (see ref. 5; $E = 1.81 \pm 0.07$; $BC = 1.20 \pm 0.19$). The proportion of variance due to environment was estimated to be the square of the ratio of the standard deviations (E/BC), for log-mass, solids and *p*H.

Fig. 2 Distribution of percentage of recurrent parent (E) genotype in the 237 back-cross progeny, estimated on the basis of the marker genotypes and their relative distances. Determination of marker genotypes was as previously described[13]. Estimates of the percentage of recurrent parent genome were produced by the recently developed computer program HyperGene™ (N. D. Young, A.H.P. and S.D.T., unpublished results). Although the average agreed closely with the Mendelian expectation of 75% for a back-cross, values for individual plants ranged from 59% to over 90%. The distribution of the proportion of recurrent-parent genome agrees with the mathematical expectation[35,36]. The individual with >90% E appears to carry only five fragments from CL (ranging from 9 to 47 map units in length) and could be returned to essentially 100% *E* with two additional back-crosses of fewer than 100 plants each, or one additional back-cross of about 550 plants. This is far more rapid than the 6–8 back-crosses routinely used to eliminate donor genome in the absence of markers.



abundant in 12 cases, whereas in nine cases the homozygote (E/E) was favoured. Overall, the effects of skewing approximately cancelled each other out: on average, the back-cross contained the expected 75% E genome (Fig. 2).

We then turned to the question of mapping the Mendelian factors that underly continuous variation in fruit mass, soluble-solids concentration and *p*H. The method of maximum likelihood and lod scores, commonly used in human linkage analysis[17], has recently been adapted[16] to allow interval mapping of QTLs. At each position in the genome, one computes the 'most likely' phenotypic effect of a putative QTL affecting a trait (the effect which maximizes the likelihood of the observed data arising) and the odds ratio (the chance that the data would arise from a QTL with this effect divided by the chance that it would arise given no linked QTL). The lod score, defined as the $\log_{10}$ of the odds ratio, summarizes the strength of evidence in favour of the existence of a QTL with this effect at this position; if the lod score exceeds a pre-determined threshold, the presence of a QTL is inferred. The traditional approach [8,9] to mapping QTLs involves standard linear regression, which accurately measures the effect of QTLs falling at marker loci only, underestimating the effects of other loci in proportion to the amount of recombination between marker and QTL. In contrast, interval mapping allows inference about points throughout the entire genome and avoids confounding phenotypic effects with recombination, by using information from flanking genetic markers. In the special case when a QTL falls exactly at a marker locus, interval mapping reduces to linear regression. A computer program, MAPMAKER-QTL, was written (S.E.L. and E.S.L., unpublished) to implement interval mapping.

Due to the large number of markers tested, an extremely high lod score threshold must be adopted to avoid false positives. Given the genetic length of the tomato genome and the density of markers used, a threshold of 2.4 gives a probability of less than 5% that even a single false positive will occur anywhere in the genome[16]. This is approximately equivalent to requiring the significance level for any single test to be 0.001.

QTL likelihood maps, showing how lod scores for fruit mass, soluble-solids concentration and *p*H change as one moves along the genome, reveal multiple QTLs for each trait and estimate their location to within 20–30 cM (Fig. 3).

(1) Factors for fruit mass were found on six chromosomes (1, 4, 6, 7, 9 and 11). In each case, CL alleles decrease fruit mass (by 3.5 to 6.0 g), adding to a total reduction of 28.1 g inferred for back-cross progeny carrying a CL allele at all six loci. This accounts for about half of the approximately 60 g difference between E and CL.

(2) Factors for soluble-solids concentration were found on

four chromosomes (3, 4, 6 and 7). In each case, CL alleles elevate soluble-solids concentration (by 0.83 to 1.89 °Brix), adding to a total of 4.57 °Brix (versus a difference of ~5 °Brix between the parental strains). This large effect in the back-cross is consistent with previous reports that high soluble-solids concentration exhibits dominance[12] and overdominance[13]. The QTL alleles for both fruit mass and soluble-solids concentration all produce effects in the direction predicted by the difference between the parental strains.

(3) Factors for *p*H were found on five chromosomes (3, 6, 7, 8 and 10). In addition, the lod score for a putative QTL on chromosome 9 fell just below our threshold. Because the parental strains do not differ greatly in *p*H, we suspected that CL alleles might not all produce effects in the same direction. In fact, *p*H was increased by four QTLs and decreased by two, including the likely QTL on chromosome 9. This provides a genetic explanation for the observation that many back-cross progeny exhibited more extreme phenotypes than the parental strains (Fig. 1), a phenomenon known as transgression[18].

Together, the QTLs identified for fruit mass, soluble solids and *p*H account for 58%, 44% and 48%, respectively, of the phenotypic variance among the back-cross progeny, with another 13%, 9% and 11% attributable to environment.

The numbers of QTLs reported for each trait must be considered a minimum estimate. Because an extremely stringent threshold was used to avoid any false positives, some sub-threshold effects probably represent real QTLs. For example, the regions near *TG19* on chromosome 1, *CD41* on chromosome 5 and *TG68* on chromosome 12 may affect soluble-solids concentration and merit further attention in larger populations. Similarly, the region near the *u* locus on chromosome 10 may contain an additional QTL affecting *p*H (see Fig. 3 legend). Moreover, we cannot rule out the presence of many additional QTLs with tiny phenotypic effects—postulated in evolutionary theory[19] and supported by some experimental evidence[20]. Also, it is conceivable that some of our apparent QTLs actually represent several closely-linked QTLs, each with small phenotypic effects in the same direction—a phenomenon that might arise particularly in regions of genetic map compression. Finally, we should emphasize that the QTL mapping here strictly applies only to the specific environment tested and to heterozygosity for CL alleles. In principle, homozygosity for CL alleles could have been studied by using an F_2 self between E and CL, but in practice too many of the progeny are sterile.

Some regions of the genome clearly exert effects on more than one trait (for example, chromosome 6; Fig. 3), providing a genetic explanation for at least some of the correlation between the traits. Although the present data are insufficient to distinguish between pleiotropic effects of a single gene and independent effects of tightly-linked loci, the frequent coincidence of QTL locations for different traits makes it likely that at least some of the effects are due to pleiotropy.

The region near *sp* on chromosome 6 has the largest effects on soluble solids and *p*H, as well as a substantial effect on fruit mass. The *sp* gene affects plant-growth habit: the dominant CL allele causes continuous apical growth (indeterminate habit), whereas the recessive E allele causes termination in an inflorescence ('determinate' or 'self-pruning' habit)[21]. Although indeterminacy has been reported previously[22] to elevate both fruit mass and soluble-solids concentration within *L. esculentum*, we associated it with reduced fruit mass in both E × CL and another interspecific cross (E × *L. cheesmanii*; A.H.P., S. Damon, J.D.H. and S.D.T., unpublished data). These differing results might be due to a second, tightly-linked locus or to unlinked modifier genes.

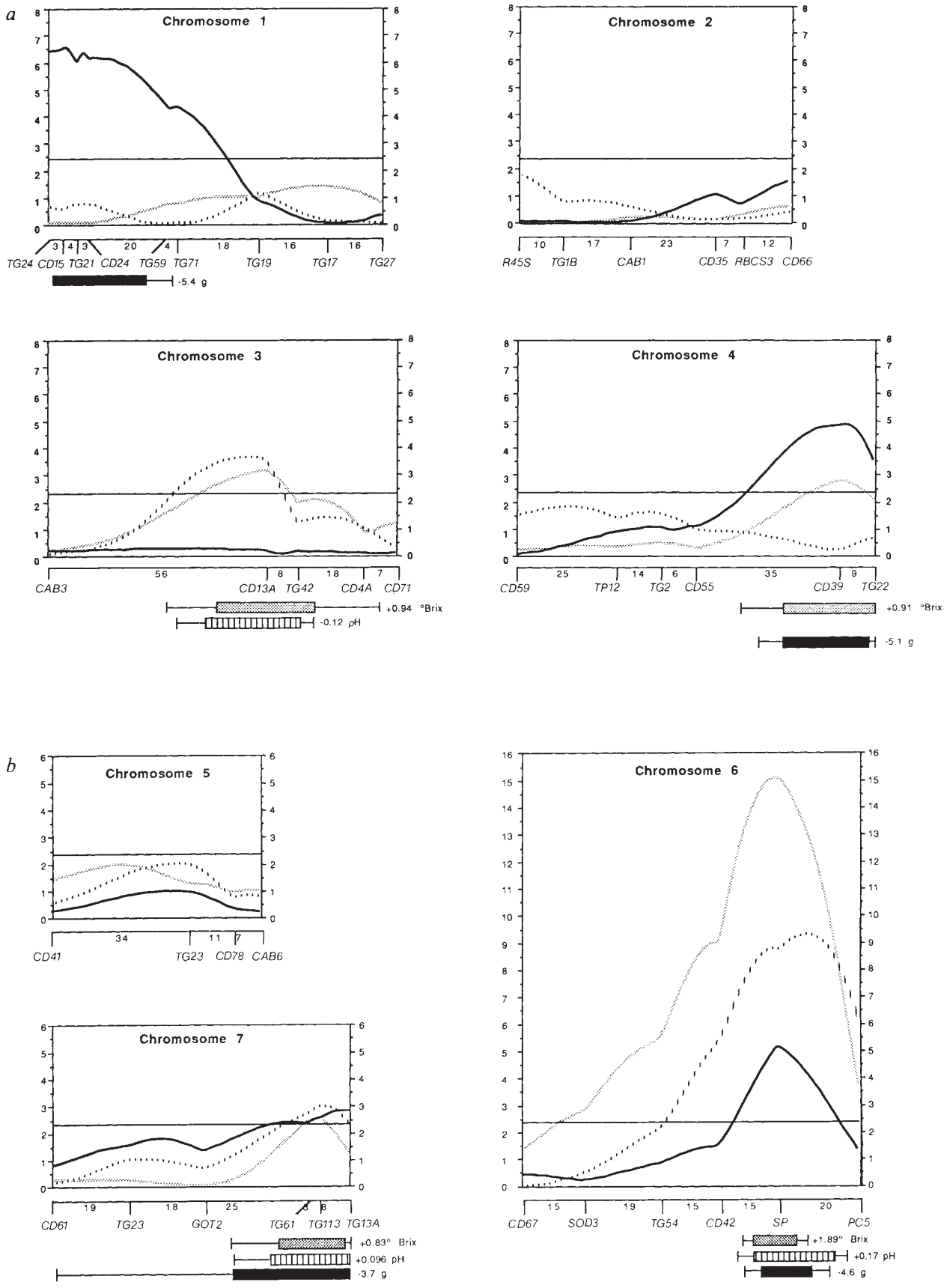
Overall, pairwise epistatic interactions between intervals were not common (about 5% of two-way analysis-of-variance tests were significant at 0.05). An interesting exception was the region near *TG16* on chromosome 8, at which the CL allele significantly enhanced the effect of three of the four QTLs for soluble-solids concentration. *TG16* also showed the most extreme segregation distortion of any marker scored (about 4:1 in favour of the E/E homozygote) and is in a region known to exhibit skewed segregation in back-crosses to other green-fruited tomato species[23,24]. The unusual properties of this region of CL clearly merit further study.

The QTLs identified here may well differ from those that would be fixed by repeated back-crossing with continuing selection for a trait, a classical method for introgressing quantitative traits. Work on LA1563, a strain with increased soluble solids produced[12] through back-crossing a different strain of E to CL has provided some suggestive evidence. By surveying RFLPs, Tanksley and Hewitt[13] recently found that LA1563 has maintained three separate regions from CL: near *CD56* on chromosome 10, near *Got2* on chromosome 7 and near *TG13* on chromosome 7. Here, we detected above-threshold effects in the last of these three regions only (which, interestingly, failed to show effects on soluble solids in a single-environment test by Tanksley and Hewitt[13]). Moreover, we detected QTLs affecting soluble-solids concentration in regions that did *not* seem to be retained. Unfortunately, the results of the two experiments are not directly comparable due to the use of a different E strain by Rick, possible environmental differences between the experiments, the possibility that small CL fragments containing QTLs went undetected in LA1563, the possibility that the region near *TG13* retained in LA1563 may not contain the QTL we detected here and the possibility that some of our sub-threshold effects are real. Although more detailed studies are clearly needed, it is interesting to speculate about why repeated back-crossing may fix a narrower class of QTLs than found by QTL mapping. Because such breeding programs[12] demand horticultural acceptability, they are likely to select against otherwise-desirable QTLs which are closely linked to undesirable effects from the wild parent. If such QTLs can first be identified by mapping, it may be feasible to remove linked deleterious effects by recombination.

Having mapped several QTLs with relatively large effects, we are now making crosses to isolate them in near-isogenic lines. These lines will be used to characterize the QTLs in various dosages, genetic backgrounds, environments and combinations. By re-assembling selected CL alleles in an otherwise E genotype, we hope to engineer an agriculturally-useful tomato with a higher yield of soluble solids.

The general approach of QTL mapping is broadly applicable to a wide range of biological endeavours. In agriculture, it might be desirable to transfer to domestic strains many quantitative traits harboured in wild species, including resistance to diseases and pests, tolerance to drought, heat, cold and other adverse conditions, efficient use of resources and high nutritional quality[25,26]. In mammalian physiology, selective breeding has generated rodent strains which differ greatly in quantitative traits such as hypertension, atherosclerosis, diabetes, predispositions to cancer, drug sensitivities and various behavioural patterns; information on the number, location and nature of these QTLs would be of value in medicine[16,27]. In evolutionary biology, the process of speciation can be investigated by studying the number and nature of genes underlying reproductive isolation[28].

The availability of detailed RFLP linkage maps[14,29-34] makes it possible to dissect quantitative traits into discrete genetic factors (QTLs): all regions of a genome can be assayed and accurate estimates of phenotypic effects and genetic position derived from interval analysis[16]. Once QTLs are mapped, RFLP markers permit genetic manipulations such as rapid construction of near-isogenic lines: flanking markers may be used to retain the QTL and the study of the remaining markers may be used to speed progress by identifying individuals with a fortuitously high proportion of the desired genetic background[38] (Fig. 2 legend). Using isogenic lines, the fundamental tools of genetics and molecular biology may be brought to bear on the study of QTLs—including testing of complementation, dominance and


a
Chromosome 1
3 4 3 20 4 18 16 16
TG24 CD15 TG21 CD24 TG59 TG71 TG19 TG17 TG27
-5.4 g
Chromosome 2
10 17 23 7 12
R45S TG1B CAB1 CD35 RBCS3 CD66
Chromosome 3
56 8 18 7
CAB3 CD13A TG42 CD4A CD71
+0.94 °Brix
-0.12 pH
Chromosome 4
25 14 6 35 9
CD59 TP12 TG2 CD55 CD39 TG22
+0.91 °Brix
-5.1 g
b
Chromosome 5
34 11 7
CD41 TG23 CD78 CAB6
Chromosome 7
19 18 25 8 8
CD61 TG23 GOT2 TG61 TG113 TG13A
+0.83° Brix
+0.096 pH
-3.7 g
Chromosome 6
15 19 15 15 20
CD67 SOD3 TG54 CD42 SP PC5
+1.89° Brix
+0.17 pH
-4.6 g

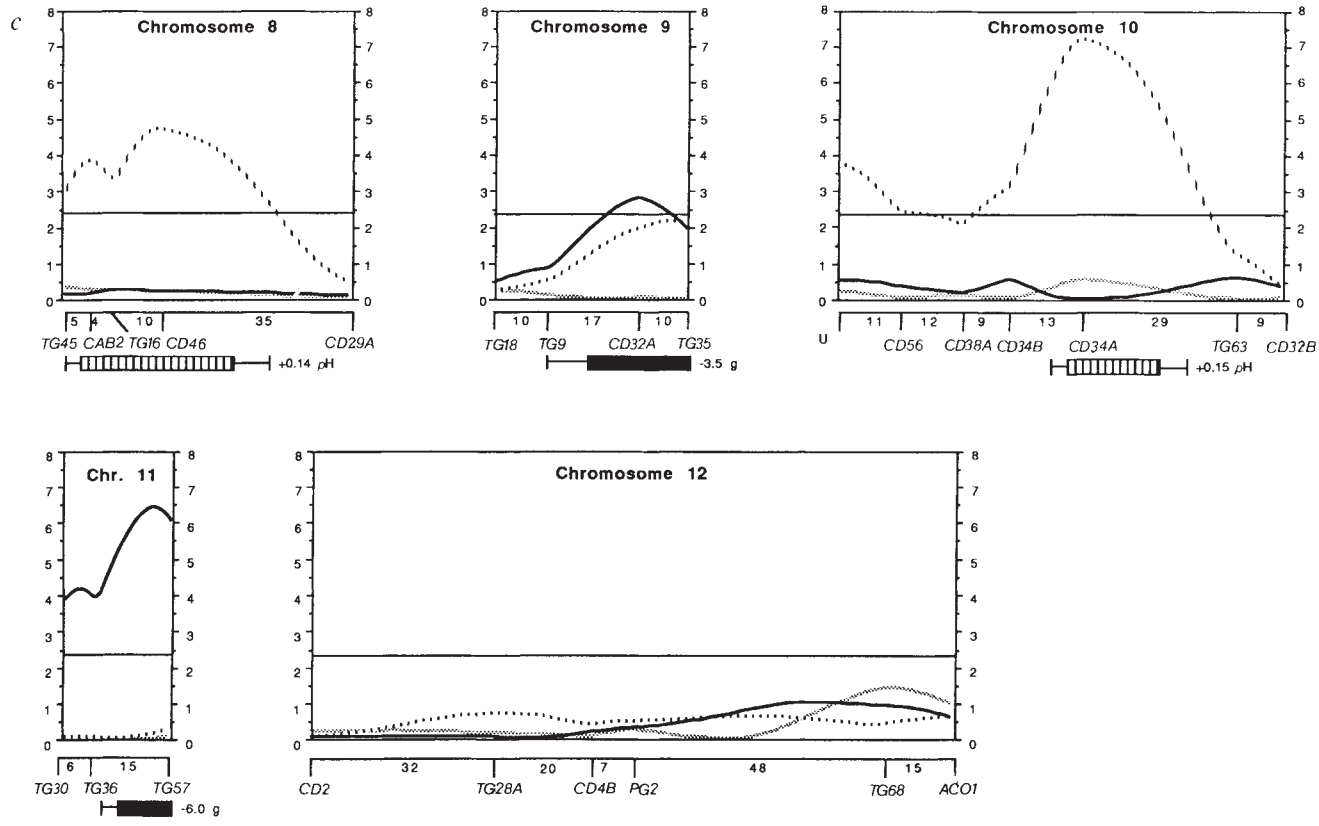


Fig. 3 QTL likelihood maps indicating lod scores for fruit mass (solid lines and bars), soluble-solids concentration (dotted lines and bars) and *p*H (hatched lines and bars), throughout the 862 cM spanned by the 70 genetic markers. The RFLP linkage map used in the analysis is presented along the abscissa, in Kosambi[37] cM. (The order of the markers agrees with the previously published map[14] of the E×P cross, except for three inversions of adjacent markers: (*TG24-CD15*), (*TG63-CD32B*) and (*TG30-TG36*). In the first case, re-analysis of the E×P data with MAPMAKER[15] indicates that the order shown here is the more likely in both E×P and E×CL. For the other two, the orders shown here are more likely in E×CL by odds of 10^4:1 and 10^7:1, but the inverse is more likely in E×P by 11:1 and 8:1 odds. These differences will be investigated in a larger E×P population.) Soluble-solids concentration and *p*H were analysed in °Brix and *p*H units, respectively; allele effects on fruit mass are presented in g; log-transformation of fruit mass was used in all analyses to achieve approximate normality. The maximum likelihood effect of a putative QTL, as well as the lod score in favour of the existence of such a QTL, have been determined at points spaced every 1 cM throughout the genome, according to Lander and Botstein[16] and a smooth curve plotted through the points. The height of the curve indicates the strength of the evidence ($\log_{10}$ of the odds ratio) for the presence of a QTL at each location—not the magnitude of the inferred allelic effect. The horizontal line at a height of 2.4 indicates the stringent threshold that the lod score must cross to allow the presence of a QTL to be inferred (see text). Information about the likely position of the QTL can also be inferred from the curve. The maximum likelihood position of the QTL is the highest point on the curve. Bars below each graph indicate a 10:1 likelihood support interval[16] for the position of the QTL (the range outside which the likelihood falls by a lod score of 1.0), whereas the lines extending out from the bars indicate a 100:1 support interval. Phenotypic effects indicated beside the bars are the inferred effect of substituting a single CL allele for one of the two E alleles at the QTL. Several regions show sub-threshold effects on one or more traits (chromosome one near *TG19*, chromosome five near *TG32* and chromosome 12 near *TG68*) which may represent QTLs but this requires additional testing. The region near *TG68* may be particularly interesting, as it is the only instance found where the CL allele seems to decrease soluble-solids concentration (by about 0.7 °Brix). In the case of chromosome 10, the lod score for *p*H crosses the significance threshold in two places. Controlling for the presence of a QTL near *CD34A*, we tested for the presence of a second QTL near *u* (by comparing the maximum lod scores assuming the presence of only the first QTL to the maximum lod score assuming the presence of two QTLs). Allowing for a QTL in the region of *CD34A*, the residual lod score near *u* falls below the required threshold. Thus, the evidence is not yet sufficient to support the presence of a QTL near *u*.

Methods. The lod score and the maximum likelihood estimate (MLE) of the phenotypic effect at any point in the genome is computed assuming that the distribution of phenotypes in the BC progeny represents a mixture of two normal distributions (of equal variance) with means depending on the genotype at a putative QTL at the given position. (Note that QTLs are considered individually and thus we did not assume that different QTL effects can be added—except in studying the possibility of two QTLs on chromosome 10 affecting *p*H.) Specifically, at a given position in the genome, the likelihood function for individual i with quantitative phenotype ϕ_i is given by $L_i(\alpha, \sigma) = (2\pi\sigma^2)^{-1/2}\{p_1 \exp(-\phi_i)^2/2\sigma^2) + p_2 \exp(-(\phi_i) - \alpha)^2/2\sigma^2\}$, where α is the effect of substituting a CL allele for an E allele at a putative QTL in the given position, σ^2 is the phenotypic variance not attributable to the QTL and p_1 and p_2 are the probabilities that individual i has genotype E/E and E/CL, respectively, at the QTL (which can be computed on the basis of the genotypes at the flanking markers and the distance to the flanking markers). The likelihood function for the entire population is $L = \Pi\, L_i$. Also α^* and σ^* denote the MLEs allowing the possibility of a QTL at the location (the values which maximize L) and σ^{**} denotes the MLE of σ, subject to the constraint that no QTL is linked ($\alpha = 0$). The lod score is then given by $\log_{10}\{L(\alpha^*, \sigma^*)/L(0, \sigma^{**})\}$. This method for QTL mapping is developed more fully in ref. 16.

epistasis; characterization of physiological and biochemical differences between isogenic lines; isolation of additional alleles by mutagenesis (at least in favourable systems); and, eventually, physical mapping and molecular cloning of genetic factors underlying quantitative traits.

We thank Janice Chen, Mark Daly, Gerald Dickinson and Mitzi Aguirre for technical assistance. We thank our colleagues and the referees for comments on the manuscript. This work was supported in part by grants from the NSF (to E.S.L. and to S.D.T.), from the System Development Foundation (to E.S.L.), and from the US Department of Agriculture (to S.D.T.).

Received 8 July; accepted 9 September 1988.

1. Mendel, G. *Verh. naturf. Ver. in Brunn.* **4**, (1868).
2. Johannsen, W. *Elemente der exakten Erblichkeitslehre* (Fischer, Jena, 1909).
3. Nilsson-Ehle, H. *Kreuzunguntersuchungen an Hafer und Weizen* (Lund, 1909).
4. East, E. M. *Genetics* **1**, 164-176 (1915).
5. Wright, S. *Evolution and the Genetics of Populations* (University of Chicago Press, 1968).
6. Sax, K. *Genetics* **8**, 552-556 (1923).
7. Thoday, J. M. *Nature* **191**, 368-369 (1961).
8. Tanksley, S. D., Medina-Filho, H. & Rick, C. *Heredity* **49**, 11-25 (1982).
9. Edwards, M. D., Stuber, C. W. & Wendel, J. F. *Genetics* **116**, 113-125 (1987).
10. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. *Am. J. hum. Genet.* **32**, 314-331 (1980).
11. Chmielewski, T. *Genet. pol.* **9**, 97-124 (1968).
12. Rick, C. M. *Hilgardia* **42**, 493-510 (1974).
13. Tanksley, S. D. & Hewitt, J. *Theor. appl. Genet.* **75**, 811-823 (1988).
14. Tanksley, S. D., Miller, J., Paterson, A. & Bernatzky, R. *Proc. 18th Stadler Genet. Symp.* (in the press).
15. Lander, E. S. *et al. Genomics* **1**, 174-181 (1987).
16. Lander, E. S. & Botstein, D. *Genetics* (in the press).
17. Ott, J. *Analysis of Human Genetic Linkage* (Johns Hopkins, Baltimore, 1985).
18. Simmonds, N. W. *Principles of Crop Improvement*, 82-85 (Longman, New York, 1981).
19. Lande, R. *Heredity* **50**, 47-65 (1983).
20. Shrimpton, A. E. & Robertson, A. *Genetics* **11**, 445-459 (1988).
21. Yeager, A. F. *J. Hered.* **18**, 263-265 (1927).
22. Emery, G. C. & Munger, H. M. *J. Am. Soc. Hort. Sci.* **95**, 410-412 (1966).
23. Zamir, D. & Tadmor, Y. *Bot. Gaz.* **147**, 355-358 (1986).
24. Tanksley, S. D. in *Isozymes in Plant Genetics and Breeding* (eds Tanksley, S. D. & Orton, T. J.) 331-338 (Elsevier, Amsterdam, 1983).
25. Rick, C. M. in *Genes, Enzymes and Populations* (ed. A. M. Srb) 255-268 (Plenum, New York, 1973).
26. Harlan, J. R. *Crop Sci.* **16**, 329-333 (1976).
27. Festing, M. F. W. *Inbred Strains in Biomedical Research* (Oxford, New York, 1979).
28. Coyne, J. A. & Charlesworth, B. *Heredity* **57**, 243-246 (1986).
29. Helentjaris, T. *Trends Genet.* **3**, 217-221 (1987).
30. Landry, B. S., Kessell, R., Leung, H. & Michelmore, R. W. *Theor. appl. Genet.* **74**, 646-653 (1987).
31. Burr, B., Burr, F. A., Thompson, K. H., Albertsen, M. C. & Stuber, C. W. *Genetics* **118**, 519-526 (1988).
32. Chang, C., Bowman, J. L., DeJohn, A. W., Lander, E. S. & Meyerowitz, E. M. *Proc. natn. Acad. Sci. U.S.A.* **68**, 6856-6860.
33. McCouch, S. R. *et al. Theor. appl. Genet.* (in the press).
34. Bonierbale, M. W., Plaisted, R. L. & Tanksley, S. D. *Genetics* (in the press).
35. Franklin, I. A. *Theor. Populat. Biol.* **11**, 60-80 (1977).
36. Stam, P. *Genet. Res.* **25**, 131-155 (1980).
37. Kosambi, D. D. *Ann. Eugen.* **12**, 172-175 (1944).
38. Tanksley, S. D. & Rick, C. M. *Theor. appl. Genet.* **57**, 161-170 (1980).

Novel anti-inflammatory peptides from the region of highest similarity between uteroglobin and lipocortin I

Lucio Miele, Eleonora Cordella-Miele, Antonio Facchiano & Anil B. Mukherjee*

Section on Developmental Genetics, Human Genetics Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, 20892, USA

Significant future developments in the effective treatment of inflammatory diseases may arise from non-toxic dual inhibitors of both cyclooxygenase and lipoxygenase pathways in the arachidonate cascade[1]. Inhibition of phospholipase A_2 (PLA_2) (EC 3.1.1.4), may provide such a dual action and recent research has concentrated on the role of PLA_2-inhibitory proteins as possible anti-inflammatory agents. Blastokinin[2] or uteroglobin[3] is a steroid-induced rabbit secretory protein with PLA_2-inhibitory activity. Its biochemical and biological properties have been extensively studied[4–14] and its crystallographic structure has been resolved at 1.34 Å (refs 15, 16). Lipocortins are a family of related proteins[17–22], which, it has been suggested, mediate the anti-inflammatory effects of glucocorticoids (for a review, see ref. 23). Some proteins of this group have been purified[24,25] and the complementary DNA sequences of two human lipocortins are known[25,26]. Lipocortins inhibit PLA_2 *in vitro*[17–21,24–29], although their mechanism of action is still unclear[24–29,30,31]. Recombinant lipocortin I inhibits eicosanoid synthesis in isolated perfused lungs from the guinea pig[32]. Here, we report that synthetic oligopeptides corresponding to a region of high amino-acid sequence similarity between uteroglobin and lipocortin I have potent PLA_2 inhibitory activity *in vitro* and striking anti-inflammatory effects *in vivo*.

Mature uteroglobin (UG) is an antiparallel dimer formed by two identical subunits of 70 amino acids each[15,16]. Figure 1*a* shows the alignment of the mature UG monomer[33] with lipocortins I and II (refs 25, 26). Lipocortins I and II are formed by four non-identical repeated units of about 70 amino acids each[34–36]. The program PRTALN aligns UG with a region of lipocortins I and II that approximately corresponds to the second repeat (Fig. 1*a*, I and III), the region of highest similarity between the two lipocortins. The total similarity (identities + conservative replacements) is 40% in both cases. However, the highest number of identities (19) is found between UG and lipocortin I repeat three (residues 199–275), with a total similarity of 43%. Moreover, a striking local similarity was identified between residues 40–46 of UG and residues 247–253 of lipocortin I, repeat three. In UG, this region corresponds to the C-terminal part of α-helix three (residues 32–47)[15,16]. As it is 'centered' on a relatively long region of local similarity, the alignment with repeat three is probably the most accurate. We drew four conclusions from these results. (1) There is amino-acid sequence similarity between UG and lipocortins; (2) UG is more similar to lipocortin I than to lipocortin II, although it can be aligned to the region of highest similarity between the two lipocortins; (3) compared to repeats two and three of lipocortin I, the highest density of identities and conservative replacements is in the C-terminal half of UG (residues 33–70) and particularly in α-helix three (Fig. 1*a*); (4) the region of most similarity between helix three of UG and repeat three of lipocortin I can be precisely identified as a heptapeptide spanning residues 40–46. This heptapeptide aligns with lipocortin I residues 247–253 (Fig. 1*a*).

Hydropathy profiles of UG and the corresponding regions of lipocortins are shown in Fig. 1*b* (I–IV). The similarity is evident over most of the UG sequence and particularly in the region of UG between positions 37–52. Interestingly, the hydropathy profile of porcine pancreatic PLA_2 (Fig. 1*b*, V) shows a striking similarity with that of the UG monomer and of the corresponding regions of lipocortins, even in the absence of significant sequence similarity. It has been independently suggested that lipocortin repeats and PLA_2 could have a similar three-dimensional organization[35]. Moreover, refined crystallographic data have shown that the molecular surface of UG is strikingly similar to the one of PLA_2 (ref. 16). These data may indicate that PLA_2, the UG monomer and the lipocortin repeats have a similar three-dimensional organization.

On the basis of computer analyses, we synthesized oligopeptides corresponding to the C-terminal part of UG α-helix three and tested them for PLA_2-inhibitory activity. Table 1 shows the amino-acid sequences and PLA_2-inhibitory properties of synthetic peptides derived from UG and lipocortin I. Peptide one (P1), which corresponds to the nine C-terminal amino-acid residues of α-helix three, is a very potent inhibitor of PLA_2, with ~80% inhibition at 50 nM, under the experimental conditions used. Peptide two (P2), corresponding to lipocortin I residues 246–254, is as active as P1 (Table 2). Peptide three (P3), which retains full inhibitory activity under these experimental conditions, was constructed by substituting an

* To whom correspondence should be addressed.