BMC Genomics

BioMed Central

Research article

Analysis of expressed sequence tags from *Actinidia*: applications of a cross species EST database for gene discovery in the areas of flavor, health, color and ripening

Ross N Crowhurst[1], Andrew P Gleave[1], Elspeth A MacRae[1], Charles Ampomah-Dwamena[1], Ross G Atkinson[1], Lesley L Beuning[1], Sean M Bulley[1], David Chagne[1], Ken B Marsh[1], Adam J Matich[1], Mirco Montefiori[1], Richard D Newcomb[1], Robert J Schaffer[1], Björn Usadel[2], Andrew C Allan[1], Helen L Boldingh[1], Judith H Bowen[1], Marcus W Davy[1], Rheinhart Eckloff[1], A Ross Ferguson[1], Lena G Fraser[1], Emma Gera[1], Roger P Hellens[1], Bart J Janssen[1], Karin Klages[1], Kim R Lo[1], Robin M MacDiarmid[1], Bhawana Nain[1], Mark A McNeilage[1], Maysoon Rassam[1], Annette C Richardson[1], Erik HA Rikkerink[1], Gavin S Ross[1], Roswitha Schröder[1], Kimberley C Snowden[1], Edwige JF Souleyre[1], Matt D Templeton[1], Eric F Walton[1], Daisy Wang[1], Mindy Y Wang[1], Yanming Y Wang[1], Marion Wood[1], Rongmei Wu[1], Yar-Khing Yauk[1] and William A Laing*[1]

Address: [1]The Horticultural and Food Research Institute of New Zealand, PB 92169, Auckland, New Zealand and [2]Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany

Email: Ross N Crowhurst - RCrowhurst@hortresearch.co.nz; Andrew P Gleave - agleave@hortresearch.co.nz; Elspeth A MacRae - Elspeth.MacRae@scionresearch.com; Charles Ampomah-Dwamena - cDwamena@hortresearch.co.nz; Ross G Atkinson - ratkinson@hortresearch.co.nz; Lesley L Beuning - lbeuning@hortresearch.co.nz; Sean M Bulley - sbulley@hortresearch.co.nz; David Chagne - dchagne@hortresearch.co.nz; Ken B Marsh - kmarsh@hortresearch.co.nz; Adam J Matich - amatich@hortresearch.co.nz; Mirco Montefiori - MMontefiori@hortresearch.co.nz; Richard D Newcomb - rnewcomb@hortresearch.co.nz; Robert J Schaffer - rschaffer@hortresearch.co.nz; Björn Usadel - usadel@mpimp-golm.mpg.de; Andrew C Allan - aallan@hortresearch.co.nz; Helen L Boldingh - hboldingh@hortresearch.co.nz; Judith H Bowen - jbowen@hortresearch.co.nz; Marcus W Davy - mdavy@hortresearch.co.nz; Rheinhart Eckloff - mwilliams@hortresearch.co.nz; A Ross Ferguson - rferguson@hortresearch.co.nz; Lena G Fraser - lfraser@hortresearch.co.nz; Emma Gera - egera@hortresearch.co.nz; Roger P Hellens - rhellens@hortresearch.co.nz; Bart J Janssen - Bjanssen@hortresearch.co.nz; Karin Klages - kklages@hortresearch.co.nz; Kim R Lo - klo@hortresearch.co.nz; Robin M MacDiarmid - rMacDiarmid@hortresearch.co.nz; Bhawana Nain - bnain@hortresearch.co.nz; Mark A McNeilage - MMcNeilage@hortresearch.co.nz; Maysoon Rassam - mrassam@hortresearch.co.nz; Annette C Richardson - ARichardson@hortresearch.co.nz; Erik HA Rikkerink - erikkerink@hortresearch.co.nz; Gavin S Ross - gross@hortresearch.co.nz; Roswitha Schröder - RSchroeder@hortresearch.co.nz; Kimberley C Snowden - KSnowden@hortresearch.co.nz; Edwige JF Souleyre - EdwigeSouleyre@hortresearch.co.nz; Matt D Templeton - Mtempleton@hortresearch.co.nz; Eric F Walton - eric.walton@otago.ac.nz; Daisy Wang - DWang@hortresearch.co.nz; Mindy Y Wang - mwang@hortresearch.co.nz; Yanming Y Wang - yanming.wang@hortresearch.co.nz; Marion Wood - mwood@hortresearch.co.nz; Rongmei Wu - rwu@hortresearch.co.nz; Yar-Khing Yauk - yyauk@hortresearch.co.nz; William A Laing* - wlaing@hortresearch.co.nz

* Corresponding author

Abstract

Background: Kiwifruit (Actinidia spp.) are a relatively new, but economically important crop grown in many different parts of the world. Commercial success is driven by the development of new cultivars with novel consumer traits including flavor, appearance, healthful components and convenience. To increase our understanding of the genetic diversity and gene-based control of these key traits in Actinidia, we have produced a collection of 132,577 expressed sequence tags (ESTs).

Results: The ESTs were derived mainly from four *Actinidia* species (*A. chinensis, A. deliciosa, A. arguta* and *A. eriantha*) and fell into 41,858 non redundant clusters (18,070 tentative consensus sequences and 23,788 EST singletons). Analysis of flavor and fragrance-related gene families (acyltransferases and carboxylesterases) and pathways (terpenoid biosynthesis) is presented in comparison with a chemical analysis of the compounds present in *Actinidia* including esters, acids, alcohols and terpenes. ESTs are identified for most genes in color pathways controlling chlorophyll degradation and carotenoid biosynthesis. In the health area, data are presented on the ESTs involved in ascorbic acid and quinic acid biosynthesis showing not only that genes for many of the steps in these pathways are represented in the database, but that genes encoding some critical steps are absent. In the convenience area, genes related to different stages of fruit softening are identified.

Conclusion: This large EST resource will allow researchers to undertake the tremendous challenge of understanding the molecular basis of genetic diversity in the *Actinidia* genus as well as provide an EST resource for comparative fruit genomics. The various bioinformatics analyses we have undertaken demonstrates the extent of coverage of ESTs for genes encoding different biochemical pathways in *Actinidia*.

Background

The genus *Actinidia* Lindl. is large, containing between 50 and 70 species of climbing plants originating mainly in southern China [1]. Over the past 30 years kiwifruit has developed into an important horticultural crop, firstly in New Zealand, and subsequently in other countries such as Chile, China and Italy [2,3]. Currently cultivars from three species are grown commercially; the green-fleshed kiwifruit, *Actinidia deliciosa* (A. Chev.) C.F.Liang et A.R.Ferguson, the closely related yellow-fleshed *A. chinensis* Planch. and the kiwiberry, *A. arguta* (Sieb. et Zucc.) Planch. ex Miq. Most of the kiwifruit cultivars grown commercially are seedling selections and there has been little systematic breeding [2]. Consequently there are still many characteristics within the genus (Fig. 1A) that could be incorporated into commercial cultivars [1] and to do this efficiently requires a better knowledge of how these characteristics are regulated. For the consumer, critical desirable attributes of kiwifruit are flavor and fragrance, appearance, healthful components and convenience.

Flavor and fragrance are determined by the chemical composition of the fruit. A wide range of compounds has been detected within the *Actinidia* genus leading to distinct and different flavors in the fruit [4]. These compounds include polyphenolics, acids, alcohols and volatile compounds such as terpenes and esters. Flavor is also influenced by the sugar to acid balance, with some growers currently receiving a premium for sweeter fruit. *Actinidia* fruit color

ranges from the original green kiwifruit, newer yellow varieties as well as red, purple and orange fruit (Fig. 1B). The extensive range of colors is caused by the presence or absence of chlorophyll, anthocyanins and carotenoids [4]. Healthful attributes of kiwifruit include its high ascorbic acid levels [2], quinic acid levels [5], and the presence of triterpenes and folic acid. On the other hand, allergens are identified as undesirable health attributes in kiwifruit and in many other fruit [4]. Convenience includes such factors as the ability to store the fruit for a long time [2] and a long shelf life, as well as easy determination of eating ripeness and having an edible or peelable skin. Little is currently known about the genetic diversity and gene-based control of these major attributes in kiwifruit.

EST databases as resources for uncovering genetic diversity have been established for many plant species (Additional file 1). Most comprehensively surveyed are *Arabidopsis* (*Arabidopsis thaliana*; 1,279,945 in GenBank) and rice (*Oryza sativa*; 1,211,418 in GenBank), both of which have been fully genome-sequenced. Fruit crops have been less extensively surveyed, but recently there have been reports on EST projects from fruit, including tomato (*Lycopersicon esculentum* [6], 258,408 sequences in GenBank), grape (*Vitis vinifera*, which has also had its genome fully sequenced [7]; [8,9], 363,365 ESTs in GenBank), apple (*Malus × domestica* [10], 255,103 sequences in GenBank) and pineapple (*Ananas comosus* [11], 5649 ESTs in GenBank). However, there are few ESTs sequenced from the

Figure 1
Fruit diversity in the genus *Actinidia*. Fruit of species used to make EST libraries are identified by letters. A is *A. arguta*, C is *A. chinensis*, D is *A. deliciosa*, E is *A. eriantha*, I is *A. indochinensis*, P is *A. polygama* and S is *A. setosa*. A. *hemsleyana* is not in the photos.

Ericales (the order to which *Actinidia* belongs) registered in GenBank (Additional file 1). These significant fruit EST resources have been used to identify genes likely to be involved in the ripening process in tomato [6] and in the generation of aroma in apple [12]. In addition, ESTs are useful sources of simple sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs), both of which are useful markers for creating genetic maps in plants [13-19].

Knowledge of the *Actinidia* genome or transcriptome is currently restricted to the 511 sequences available in GenBank (dbEST, January 2008). To improve and develop new varieties of kiwifruit with the desired flavor, health and convenience attributes, it is useful to use modern genomics techniques in conjunction with breeding tools such as marker-assisted selection and genetic transformation. For this reason, we undertook a major EST sequencing project in *Actinidia* to develop a basic genetic resource covering a range of different species and tissues. In this

paper we analyze more than 130,000 ESTs, derived mainly from four species in the *Actinidia* genus (Table 1). This information is presented with reference to the chemical composition of these species and using specific examples where we have increased our understanding of the genetic diversity and gene-based control of critical attributes.

Results
Overview
Actinidia cDNA libraries were constructed mainly from four *Actinidia* species: *A. deliciosa*, *A. chinensis*, *A. eriantha* and *A. arguta* (Table 1, Additional file 2). The libraries were constructed from petals, fruits, buds and leaves, with a small number from roots and cell culture and sequenced from the 5' end. The average edited sequence length of the 132,577 ESTs was 503 bases. Clustering these sequences using a 95% threshold resulted in 18,070 tentative consensus (TC) sequences (average length 577 bases), with 23,788 ESTs remaining as singletons. The combination of TCs and singletons are referred to as non redundant (NR) clusters. These numbers are similar to those observed with apple [10]. Over 95% of TCs had fewer than 18 EST members, and the largest single TC had 758 members. It should be noted that *A. chinensis* and *A. deliciosa* are more closely related to each other than the other species mentioned in this paper, and that only a restricted number of genotypes within each species was used in making EST libraries compared to the total numbers available.

Sequence analysis
Codon usage
Knowledge of the GC content of a genome and codon usage is useful when devising PCR-based strategies for mapping and gene isolation, as well as for hybridization studies by microarray. Based on the coding regions of 302 *A. deliciosa*, 319 *A. chinensis* and 84 *A. eriantha* full-length cDNAs, the GC content in the third base position was estimated to be 55%, 49% and 58%, respectively. These values are higher than the overall GC ratio from the sequences of the complete EST datasets of *A. deliciosa* (46%), *A. chinensis* (46%) and *A. eriantha* (48%), indicating that there is some pressure, particularly in *A. deliciosa* and *A. eriantha*, towards an increased GC ratio in coding regions compared with non-translated regions. Overall the codon usage of the three *Actinidia* species is similar, although not identical (Additional file 3). *A. deliciosa* and *A. eriantha* differ in only their preference for aspartate (GAT/GAC) and serine (TCT/TCC). In contrast, comparative codon preferences between *A. deliciosa* and *A. chinensis* show differences for alanine (GCC/GCT), glycine (GGC/GGA), isoleucine (ATC/ATT), leucine (CTC/TTG), asparagine (AAC/AAT), glutamine (CAG/CAA), threonine (ACC/ACT) and valine (GTG/GTT). In most of these codons, *A. chinensis* shows a greater preference for an A or

Table 1: Numbers of ESTs obtained from different *Actinidia* tissues by species.

	Species					
Tissue	*A. arguta*	*A. chinensis*	*A. deliciosa*	*A. eriantha*	Other[a]	Total[b]
Bud	c	15,689	3,4519			50,208
Cell[d]		4,851				4,851
Fruit	5,421	8,453	13,282	11,259		38,415
Leaf		17,325				17,325
Petal	1,836	1,061	9,950	1,388	1,422	15,657
Root					5,101	5,101
Stem					1,020	1,020
Total	7,257	47,379	57,751	12,647	7,543	132,577

The number of ESTs was calculated by amalgamating libraries from the same tissues.

[a] Other miscellaneous species were *Actinidia setosa* (1,020 from stem), *A. hemsleyana* (5,101 from roots), *A. polygama* (1,348 from petal) and *A. indochinensis* (74 from petal).

[b] See Additional file 2a for full details of libraries and tissues.

[c] Empty cells had no ESTs sequenced.

[d] Cell refers to cell culture.

T in the third base position, accounting for its lower GC third base percentage when compared with the other two species.

Polymorphisms and genetic markers
Both SNPs and microsatellites or SSRs are valuable tools for genetic mapping within breeding populations of many crops. A total of 32,764 biallelic SNPs were detected from the overlapping regions of 3,901 (21.6%) of the 18,070 TCs, at a rate of one SNP every 417 bp. This frequency is higher than that reported in apple using a dataset of comparable size (one SNP per 706 bp from ~150,000 ESTs [10]). As several *Actinidia* species were used to construct the cDNA libraries, while the apple cDNA libraries were from different varieties of the same species, this increased frequency of SNPs could reflect the greater genetic diversity in the surveyed *Actinidia* EST libraries. For example the frequency in apple increased significantly to 1 in 149 bp when a greater diversity of genotypes was added (Chagne *et al* unpublished results). The polyploidy nature of several of the *Actinidia* species used may also have affected SNP frequencies. As a result of this, some of the SNPs identified may not be allelic in nature but due to homoeologous or paralogous sequences clustering in the same TC.

EST-derived SSRs have already been shown to be valuable mapping tools in *Actinidia chinensis* [14] where 150 SSRs with more than 10 dinucleotide repeat units were tested as markers in an *A. chinensis* mapping population. More than 90% of the SSR markers were polymorphic and segregated within the population. Subsequently, 20 of these SSRs were shown to be transportable across multiple species of *Actinidia*, showing the value of this resource [15].

Compared with apple [10], where less than 20% of the NRs contained a microsatellite, over 30% of *Actinidia* NRs were found to have at least one. The other major difference between apple and *Actinidia* was that, while di- and tri-nucleotide repeats were equally frequent in apple at 7 to 8% frequency, di-nucleotide repeats (18%) were twice as frequent as tri-nucleotide repeats (9%) in *Actinidia*. Tetra-nucleotide repeats were found at similar frequencies in NRs for the two genera. The lengths of the SSRs were similar in both apple and *Actinidia*, with 50% of the repeats having between 12 and 14 bases. As in apple, AG repeats were the most frequent (16% of NRs, significantly more frequent than in apple) and were followed by AT (1.2%) and AC (0.8%) while GC repeats were very infrequent.

The position of SSRs in relation to the putative initiation ATG was very similar when comparing *Actinidia* with other plants such as *Arabidopsis* and apple [10,20], with di-nucleotide repeats being predominantly (93%) in the 5' UTR, and tri-nucleotide repeats being more evenly distributed along the gene (37%).

Comparative genomic DNA hybridizations by microarray
A 17,472 feature *Actinidia* oligonucleotide microarray was constructed based on sequence data from the *Actinidia* EST database. The 44–55 mer oligonucleotides were derived from gene sequences of different *Actinidia* species (Additional file 4). To test the cross hybridization between the different genomes of the *Actinidia* species, genomic DNA from two *Actinidia* species (*A. deliciosa* and *A. eriantha*) was hybridized to two different microarray slides. A plot of average signal intensity for the two species against each other is shown in Fig. 2. Of the 13,443
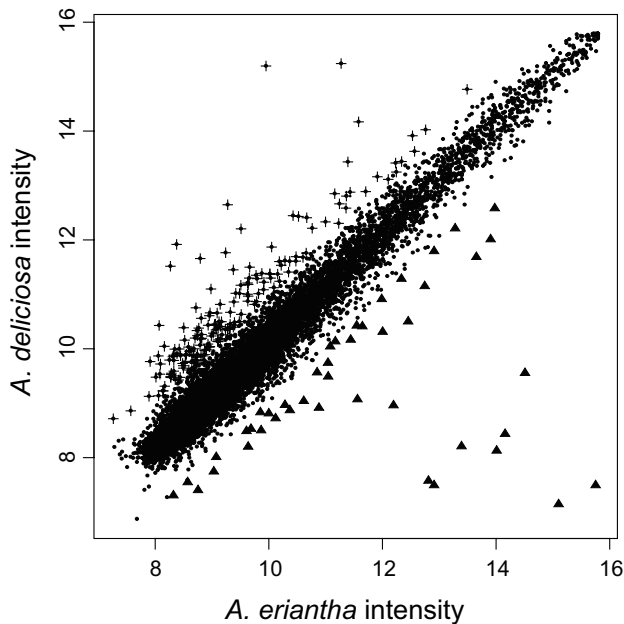
Figure 2
Plot of *Actinidia deliciosa* versus *A. eriantha* microarray intensities. Genomic DNA was labeled with cy3 and cy5 in two separate reactions, combined and hybridized to a single microarray slide. Spots not found or flagged "bad" were removed and each channel was normalized using a quantile normalization. Of the two arrays, 15,024 were flagged "good" using *A. deliciosa* DNA and 13,785 were flagged "good" using *A. eriantha* DNA. Of these spots, 13,443 were common to both species. Black spots represent oligonucleotides that showed similar intensity of hybridization between the two species. Triangles represent oligonucleotides that showed an increase in intensity with *A. eriantha* genomic DNA and crosses represent oligonucleotides that showed an increase using *A. deliciosa* genomic DNA.

informative features, the majority (98.4%) showed a similar level of hybridization between the two species (less than an average of 2-fold change), with 172 showing a higher signal intensity in *A. deliciosa* (plus symbols, Fig. 2), and 44 showing a higher intensity in *A. eriantha* (trian-

gles, Fig. 2). There was no correlation between the species from which the oligonucleotide was derived and the increased or decreased level of hybridization (Table 2). Consequently, the increases are probably due to separate gene duplications in each species.

Functional analysis
Highly populated tentative consensus sequences
The TC with the greatest number of ESTs was a cysteine proteinase (actinidin family, 758 members) followed by a metallothionein (547 members) (Additional file 5). However, when similar proteins were taken into account (i.e. different TCs that all matched a common *Arabidopsis* protein as their closest homolog with E < $1.0e^{-100}$), there were 1,266 ESTs matching cysteine proteinases followed by 769 metallothionein protein-encoding ESTs.

It was not surprising to identify the cysteine protease actinidin as the most highly abundant TC cluster in *Actinidia*. Actinidin can constitute up to 50% of soluble protein in mature *A. deliciosa* fruit at harvest [21]. Recent research has shown that this cysteine protease exists in basic and acidic forms and the amount of each form varies between species, with fruit of the important *A. chinensis* cultivar 'Hort16A' being almost devoid of acidic actinidin [22]. ESTs for actinidin occur throughout the fruit libraries that make up the database, and these cluster into 10 distinct genes/alleles [22]. As two of the key fruit libraries were subtracted for actinidin, the number of ESTs in the TC is likely to be an under-representation of the actual number. The function of this protease may be related to insect defense [23].

The metallothionein TC cluster contains a gene (pKIWI504) previously identified in kiwifruit [24] as being highly expressed in young fruit with reduced expression in the later stages of fruit development. However, the *in vivo* function of the metallothionein TC cluster is currently unknown. Interestingly, pineapple also had a high proportion of metallothionein ESTs [11].

Table 2: Comparative genomic hybridizations of different *Actinidia* species.

	A. arguta	*A. chinensis*	*A. deliciosa*	*A. eriantha*	*A. hemsleyana*	*A. indo-chinensis*	*A. poly-gama*	*A. setosa*
No. of oligos[a]	479[b]	8,957	6,606	978	148	12	188	104
A	2	58	95	15	0	0	1	1
B	0	12	27	5	0	0	0	0

A # features where the signal is increased in *A. deliciosa*.
B Signal increased in *A. eriantha*.
[a] The number of oligonucleotide primers on the microarray derived from each Actinidia species.
[b] Oligonucleotide microarrays were hybridized with labeled genomic DNA from A. deliciosa or A. eriantha. The signals were normalized and selected as differentially changing between the two species using a threshold of two.

Table 3: Functional classification of ESTs and NRs from *Actinidia*.

Bin Code	Bin Name	% *Actinidia* ESTs[a]	% *Actinidia* NR sequences[b]	% *Arabidopsis* sequences	% Potato NRs
1	PS	2.18	1.22	0.73	0.75
2	major CHO metabolism	0.86	0.72	0.37	0.54
3	minor CHO metabolism	1.02	0.75	0.45	0.41
4	glycolysis	1.03	0.60	0.23	0.27
5	fermentation	0.38	0.21	0.05	0.07
6	Gluconeogenesis/glyoxylate cycle	0.14	0.11	0.04	0.02
7	OPP	0.25	0.22	0.11	0.12
8	TCA/org. transformation	0.64	0.55	0.27	0.27
9	mitochondrial electron transport/ATP synthesis	0.82	0.71	0.49	0.34
10	cell wall	3.20	2.27	1.90	1.11
11	lipid metabolism	1.78	1.91	1.49	1.32
12	N-metabolism	0.08	0.16	0.09	0.08
13	amino acid metabolism	2.64	2.00	1.07	1.18
14	S-assimilation	0.10	0.08	0.05	0.05
15	metal handling	2.53	0.59	0.34	0.20
16	secondary metabolism	3.02	2.09	1.64	1.42
17	hormone metabolism	2.16	1.99	2.27	1.69
18	co-factor and vitamin metabolism	0.10	0.14	0.16	0.15
19	tetrapyrrole synthesis	0.28	0.26	0.17	0.16
20	stress	4.08	3.33	3.48	2.64
21	redox. regulation	1.02	1.00	0.70	0.64
22	polyamine metabolism	0.29	0.12	0.05	0.07
23	nucleotide metabolism	0.59	0.68	0.57	0.49
24	biodegradation of xenobiotics	0.04	0.08	0.09	0.06
25	C1-metabolism	0.19	0.19	0.14	0.12
26	miscellaneous	4.72	4.59	5.38	3.69
27	RNA	7.86	9.86	10.88	6.82
28	DNA	1.60	2.05	1.99	1.25
29	protein	18.61	16.94	13.31	9.59
30	signalling	3.42	4.77	4.46	3.10
31	cell	3.23	3.04	2.57	1.68
33	development	1.29	1.58	2.24	1.33
34	transport	3.62	3.99	3.49	2.58
35	not assigned	26.23	31.21	38.72	55.79

[a] Actinidia NRs were assigned to a Mapman (Thimm et al., 2004) bin according to their nearest Arabidopsis BLAST hit. ESTs belonging to the Actinidia NR were then assigned to the same bin.
[b] Potato and Arabidopsis functional assignments were as described [27,100]

Fourteen TCs in *Actinidia* with their proteins annotated by BiolView as "not assigned-unknown" proteins (i.e. matches to sequences with unknown function from *Arabidopsis* or other species) were highly expressed with over 100 EST members. Four of the unassigned TCs contained more than 300 EST members. These four TCs included kiwellin [25], which is a homolog to a grape ripening-related protein [26] with no *Arabidopsis* homologs (527 EST members); a homolog of At5g11420, a protein of unknown function (339 EST members); a senescence-related protein (no *Arabidopsis* matches, 311 EST members); and an ABA stress-related protein (no *Arabidopsis* matches, 302 EST members).

Functional analysis using Mapman and InterPro
Functional analysis of ESTs and NRs using Mapman [27] is shown in Table 3. The analysis was conducted by iden-

tifying each NR's nearest homolog to *Arabidopsis,* using BLASTx and identifying domains/families that are indicative for a given function using InterProScan. Then that NR and its member ESTs were assigned to the corresponding bin of the *Arabidopsis* match. Of the 41,858 NRs, 28,345 had sufficient homology to an *Arabidopsis* sequence to assign it a functional classification in Mapman. Thus, 32% of NRs from *Actinidia* had no *Arabidopsis* homolog, but a proportion would have homologs in other crops (e.g. genes related to fruit ripening).

A comparison of the percentage of NRs assigned to the top level Mapman classification bins and the percentage of *Arabidopsis* genes classified in the same bin showed a strong relationship with an r^2 value of 0.96 and a slope of 0.87, near to the expected 1:1 slope. The same relationship when calculated for second level bins gave an r^2 of

0.98 and a slope of 1.14 (second level bin data not shown). Taking only those top level bins that contained more than 100 NRs to remove chance as a major factor, only two bins appeared to be over-represented (i.e. > 2.0 times higher than the mean) in *Actinidia*, namely bin 4 (glycolysis) and bin 8 (TCA cycle). This may reflect a high level of primary metabolism in fruit. Interestingly, when all NRs are included, bin 5 (fermentation) and bin 6 (gluconeogenesis) were added to the list of over-represented NRs. The same relationships for potato showed no anomalies (Table 3). This analysis demonstrates that the clustering of ESTs generated a fair representation of the distribution of genes.

Overall, the average number of ESTs per NR in the database was 4.0. There were only two bins with a high number of ESTs per NR, bin 15 (metal handling) with over 17 ESTs per NR and bin 22 (polyamine metabolism) with over 9. However, bin 22 had a low number of NR and EST members, suggesting this result may have occurred by chance. In contrast, bin 15 included the metallothioneins and the very highly expressed EST discussed earlier (Additional file 5).

TCs with high numbers of ESTs (> 100) came from 18 of the 35 Mapman bins [27] with approximately three quarters falling into the bin names "not assigned, no ontology", "protein degradation", "metal handling binding chelation and storage", "stress biotic", "cell wall degradation" and "secondary metabolism isoprenoids".

The most common InterPro families [28] were also analyzed (Additional file 6). There were 3,111 InterPro families (flagged as true by InterPro) represented in the *Actinidia* NRs, with the most NRs found in the protein kinases grouping. Following that were thioredoxins, while cysteine proteinases, which had the highest number of ESTs in the database, were represented as the 7th most frequently occurring NR.

Genes related to flavor and fragrance
Understanding the relationship between fruit flavor and the genetic diversity present in the EST database requires a detailed analysis of the flavor and fragrance compounds present in the *Actinidia* genus. Fruit and flower samples from the major *Actinidia* varieties that were used to produce the EST database were characterized using both headspace sampling and solvent extraction followed by GC-FID/MS analysis. The results of this compound analysis are presented in Additional file 7 for esters (over 55 identified compounds), acids, alcohols and terpenoids.

Compound analysis
The most abundant volatile compounds in the flowers of these *Actinidia* species were alcohols and terpenes. *A. chin-*

ensis flowers contained 71% sesquiterpenes, largely dominated by farnesols (Additional file 7). *A. deliciosa* flowers had the highest volatile ester content at around 12%, and aromatic (6–9%) and straight-chain alcohols (9%). These flowers were notable for their α-farnesene (12–26%) and germacrene D (0.3–12%) content. *A. arguta* flowers contained branched-chain (13%) and aromatic alcohols (14–50%), and monoterpenes (12–24%), which were mostly linalool oxidation products reported previously [29]. *A. eriantha* flowers contained few volatile compounds, the notable ones being 2-phenylethanol (8.5%), 6-methyl-5-hepten-2-one (also 8.5%), and β-myrcene (5%). The flowers from two further *Actinidia* species showed distinct volatile profiles. *A. polygama* flowers contained linalool oxides (36.5%) and dehydroiridodial (37%), a monoterpene related to the *Nepeta cataria* (catnip) compounds [30,31]. The flowers of *A. indochinensis* flowers were dominated by 2-phenylethanol (25.8%) and a couple of terpenes, the major one being linalool (23.6%). The aromas from these flowers have also had descriptors attached (McNeilage et al. unpublished observations). For example *A. eriantha* has been described as coconut and berry, and *A. arguta* with descriptors including sweet, lime, rose and vanilla [29]. *A chinensis* and *A. deliciosa* are described as having tea-rose aromas, sweet, fresh, lilac and violet, while *A. polygama* has gardenia and lime descriptors. These descriptors reflect the diversity of volatile compounds found in the *Actinidia* flowers.

In fruit, esters accounted for up to 85% of the total volatiles. The main *A. chinensis* esters were the fresh, fruity ethyl butanoate which, depending upon analytical method and genotype, ranged from 3 to 54%, and butyl butanoate, both of which are common fruit esters [32]. *A. chinensis* 'Hort16A' fruit contained significant amounts of eucalyptol (0.2 to 21%) which has been described as having a fresh, pungent cooling taste [32]. Acetaldehyde, hexanal, E-2-hexenal and ethyl butanoate are known to be important contributors to Hort16A aromas [33]. Ethyl butanoate along with methyl butanoate and ethyl acetate were the major (~60%) compounds in the green *A. deliciosa* fruit and are known to be major volatile contributors to its flavor [34]. The other major compound was ethanol (16%), with only trace levels of other alcohols and terpenes. Whilst at more modest levels in *A. arguta* fruit (2–8%), ethyl butanoate was still substantial. *A. arguta* fruit were notable for their content of methyl- and ethyl benzoate, which are responsible for the distinctive aroma of feijoa (*Feijoa sellowiana* O. Berg) fruit [35], and also camphor (7%). The dominant ester in *A. eriantha* fruit was butyl acetate (12%), a flavor compound found in many fruit, which has a strong, fruity odor and a taste reminiscent of pineapple [32,36]. However, the total volatile level in *A. eriantha* is lower than other species and the number of esters is also lower (Additional file 7).

Ester biosynthesis

Esters are synthesized from their acid and alcohol precursors by acyltransferases (ATs). We identified 30 ATs in the *Actinidia EST* database, of which 25 are full length. These sequences were analyzed for their phylogenetic relationship with other plant ATs of known function (Fig. 3). Two clades were identified that contained enzymes involved in the synthesis of flavor-related esters (Clades 1 and 2 in Fig. 3). The first contained alcohol acyltransferases (AATs) from apple, melon and banana and nine AATs from *Actinidia* (AdAT1, AdAT2, AeAT9, AcAT15, AcAT16, AdAT17, AaAT18, AcAT20, AcAT23), while the second clade contained AATs from strawberry and three AATs from *Actinidia* (AdAT6, AdAT22, AcAT24). An *Actinidia* AT predicted to be involved in esterifying anthocyanins (AcAT25) and two ATs predicted to be anthranilate acyltransferases (AdAT8, AcAT21) were also identified. Three *Actinidia* ATs were closely related to the carboxylesterase-related protein 2 CER2 and therefore predicted to be elongases (AcAT30, AcAT27 and AcAT26) possibly involved in wax production.

Multiple sequences of the same genes were recovered for some genes reflecting orthologs, alleles or sequences from different genomes within polyploid genomes. For example, the three ATs, AcAT16, AdAT17 and AdAT1, and the two ATs, AaAT18 and AdAT2, are probable variants of just two different AATs. *A. deliciosa* is a hexaploid (6n) possibly explaining this multiplicity of genes. We may have uncovered at least one instance where, for a single AAT, all three genomic versions have been isolated (AcAT16, AdAT17 and AdAT1).

Carboxylesterases conduct the opposite reaction to ATs, i.e. the hydrolysis of esters into acids and alcohols [37]. As well as ester hydrolysis, plant carboxylesterases have been implicated in isoflavonoid biosynthesis [38], plant defense [39], and hormone regulation [40]. Analysis of *Actinidia* flowers and fruits showed a wide range of acids and alcohols (Additional file 7).

Previously published plant carboxylesterase (CXE) genes from GenBank were used to identify CXE genes in the *Actinidia* EST database using BLAST searches. The sequence alignment of CXEs revealed amino acid motifs and secondary structural features characteristic for members of the α/β hydrolase fold superfamily and more particularly of CXEs such as an active site serine surrounded by a GXSXG motif.

In *Actinidia* 19 CXEs were identified, 18 of which are full length (Fig. 4). These include CXE members likely to encode gibberellin receptors (AdCXE9, AdCXE17, AdCXE15), an isoflavanone dehydratase (AeCXE11) and an ortholog of a plant defense-associated CXE (AcCXE12).

Other CXEs from *Actinidia* align well with *Arabidopsis* CXEs and can be associated with the classes identified elsewhere [41]. The structure of an *Actinidia* CXE is now available (AeCXE1), the first plant CXE to have its structure solved [42]. AeCXE1 can hydrolyze a range of substrates from C2 to C16 esters, with a preference for C4 moieties. There are two CXEs from *Actinidia* (AdCXE16, AaCXE2) that may be orthologs to AeCXE1 from *A. eriantha*.

Terpene biosynthesis

Actinidia terpenoid compounds such as monoterpenes (C10), sesquiterpenes (C15) diterpenes (C20), triterpenes (C30), carotenoids, sterols, phytols and quinones are derived from two common precursor molecules, isopentenyl diphosphate and dimethylallyl diphosphate. These precursors are produced via the cytoplasmic mevalonate pathway or the chloroplastic 1-deoxy-D-xylulose-5-phosphate (mevalonate-independent) pathway [43]. Genes for enzymes in both these pathways are present in the *Actinidia* EST database, with the results for analysis of ESTs in the early steps of the mevalonate pathway shown in Fig. 5. The conversion of these precursors into other intermediates of the terpenoid biosynthetic pathway, including geranyl diphosphate, farnesyl diphosphate, and geranylgeranyl diphosphate, is carried out by polyisoprene synthase genes (EC 2.5.1.x). Over 200 ESTs with sequence homology to known polyisoprene synthase genes were found in the *Actinidia* EST database including 116 ESTs with similarity to dimethyallyltranstransferase (EC 2.5.1.1), 66 ESTs with homology to known monoterpene and sesquiterpene synthases and 38 ESTs homologous to squalene synthase (EC 2.2.1.21). No ESTs for phytoene synthase (EC 2.5.1.32) were identified.

The EST database collection has been used to identify a multifunctional terpene synthase gene from *A. deliciosa* flowers that produces the sesquiterpene germacrene D and a range of other sesquiterpene products at lower abundance [44]. This gene is represented by 10 sequences in the EST database.

Genes related to color

Chlorophyll catabolism

Green fleshed commercial kiwifruit such as *A. deliciosa* 'Hayward' are characterized by the retention of chlorophylls in the flesh of ripe fruit [45]. In contrast, the newer yellow-fleshed fruit cultivars (e.g. *A. chinensis* 'Hort16A' and 'Jintao') degrade the chlorophyll during ripening and the consequent loss of the green color uncovers the underlying yellow carotenoid pigmentation.

The degradation of chlorophylls (leading to the production of colorless breakdown products) is controlled by catabolic enzymes in the chlorophyll degradation path-
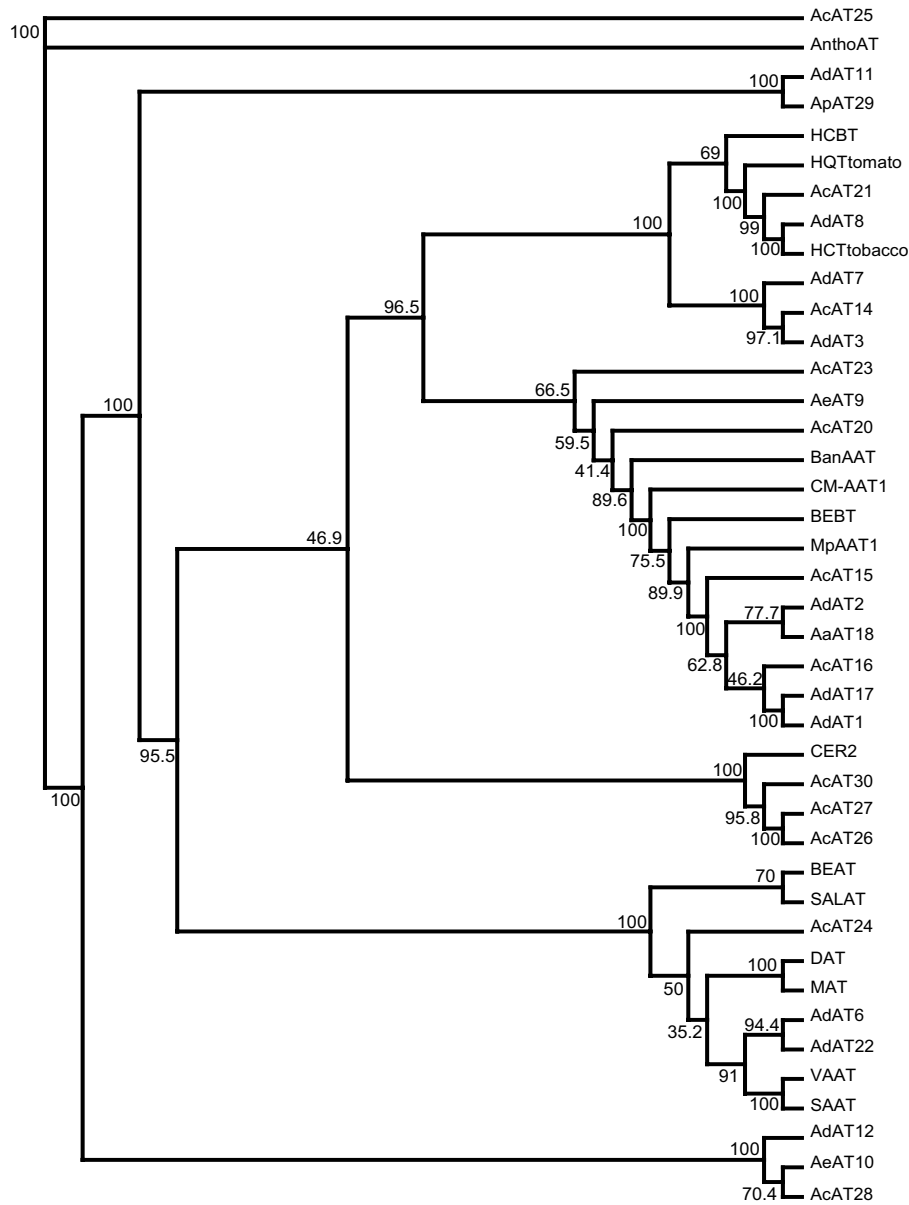
Figure 3
Phylogenetic tree of plant acyltransferases (ATs) of known function and *Actinidia* members of the acyltransferase family. Previously published plant acyltransferase sequences from GenBank were used to identify genes from the *Actinidia* EST database using BLAST searches. Abbreviations for species and AT names are as follows: AdAT, *Actinidia deliciosa* AT; AeAT, *A. eriantha* AT; AcAT, *A. chinensis* AT; AaAT, *A arguta* AT; DAT, *Catharanthus roseus* deacetylvindoline 4-0-acetyltransferase, GenBank Accession No. AF053307; MAT, *C. roseus* minovincinine 19-hydroxy-O-acetyltransferase, AAO13736; BEAT, *Clarkia breweri* acetyl-CoA:benzylalcohol acetyltransferase, AF043464; SALAT, *Papaver somniferum* salutaridinol 7-O-acetyltransferase, AF339913; BEBT, *Clarkia breweri* benzoyl-CoA:benzyl alcohol benzoyl transferase, AF500200; MpAAT1, *Malus pumila* alcohol acyltransferase; AY707098; CM-AAT1, *Cucumis melo* alcohol acyltransferase, CAA94432; SAAT, *Fragaria × ananassa* alcohol acyltransferase, AAG13130; HCBT, *Dianthus caryophyllus* anthranilate N-hydroxycinnamoyl benzoyltransferase, Z84383; AnthocyaninAT *Petunia frutescens* anthocyanin acyltransferase, BAA93453; VAAT, *Fragaria vesca* alcohol acyltransferase, AX025504; BanAAT, *Musa acuminata* alcohol acyltransferase, AX025506; CER2, *A. thaliana* CER2 gene, X93080; HCT, *Nicotiana tabacum* hydroxyl-cinnamoyl transferase, AJ507825; HQT, *Lycopersicon esculentum* hydroxycinnamoyl CoA quinate transferase, AJ582652. Percentage bootstrap values (1000 bootstrap replicates) for groupings are given by each branch. Transf. on cladogram is transferase.
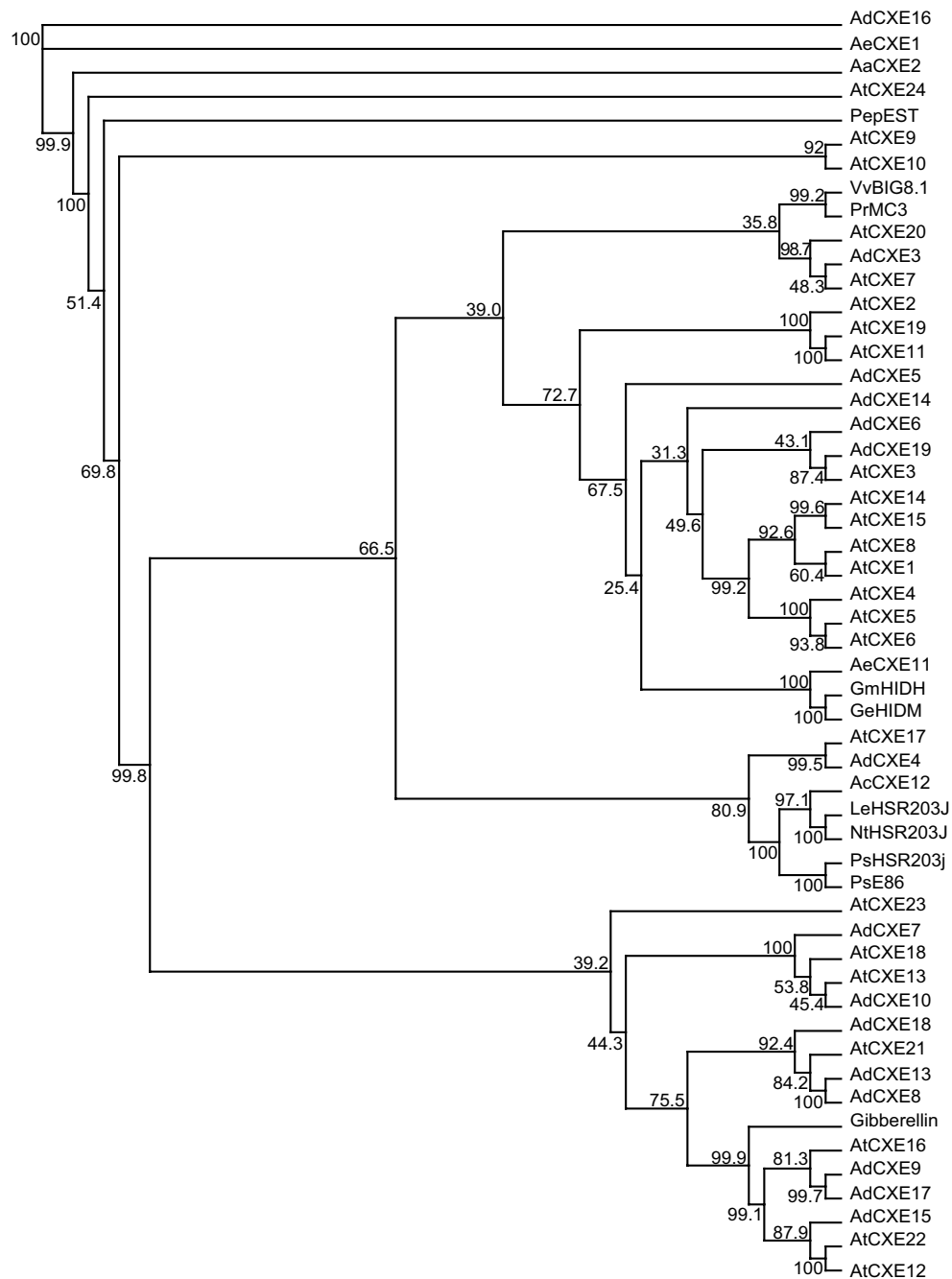
Figure 4
Phylogenetic tree of plant carboxylesterases of known function and *Actinidia* members of the carboxylesterase family (CXE).
Previously published plant carboxylesterase sequences from GenBank were used to identify genes in the *Actinidia* EST database
using BLAST searches. Abbreviations for species and CXE names are as follows: AtCXE, *Arabidopsis thaliana*, CXE; AaCXE, *A.
arguta* CXE; AcCXE, *A. chinensis* CXE; AdCXE, *A. deliciosa* CXE; AeCXE, *A. eriantha* CXE; PepEST, *Capsicum annuum* Esterase,
GenBank Accession No. AAF77578.1; Gibberellin, Gibberellin receptor GID1, Q6L545; LeHSR203J, *Lycopersicon esculentum*
HSR203J, BAA74434.1; NtHSR203J, *Nicotiana tabacum* HSR203J, AAF62404.1; PsHSR203j, *Pisum sativum* HSR203J,
BAA85654.1; PsE86, *P. sativum* E86, BAA85654.1; GeHIDM, *Glycyrrhiza echinata* HIDM, BAD80839.1; GmHIDH, *Glycine max*
HIDH, BAD80840.1; PrMC3, *Pinus radiata* MC3, AAD04946.2; VvBIG8.1, *Vitis vinifera* BIG8.1, >AF48726-1. Percentage boot-
strap values (1000 bootstrap replicates) for groupings are given by each branch.

Compounds	Enzyme	ESTs

mevalonate

↓ mevalonate kinase EC 2.7.1.36 — 8

5-phosphomevalonate

↓ phosphomevalonate kinase EC 2.7.4.2 — 2

5-diphosphomevalonate

↓ diphosphomevalonate decarboxylase EC 4.1.1.33 — 42

isopentenyl diphosphate

↕ isopentenyl-diphosphate delta-isomerase EC 5.3.3.2 — 29

dimethylallyl diphosphate

↓ dimethylallyltranstransferase EC 2.5.1.1 — 116

geranyl diphosphate ⟶ **monoterpenes**

polyisoprene synthase EC 2.5.1.x

↓

farnesyl diphosphate ⟶ **sesquiterpenes** — 66

polyisoprene synthase EC 2.5.1.x

↓

geranylgeranyl diphosphate ⟶ **diterpenes**

squalene synthase EC 2.5.1.21 — 38

↓

squalene ⟶ **triterpenes**

squalene monooxgenase EC 1.14.99.7 — 20

↓

sterols

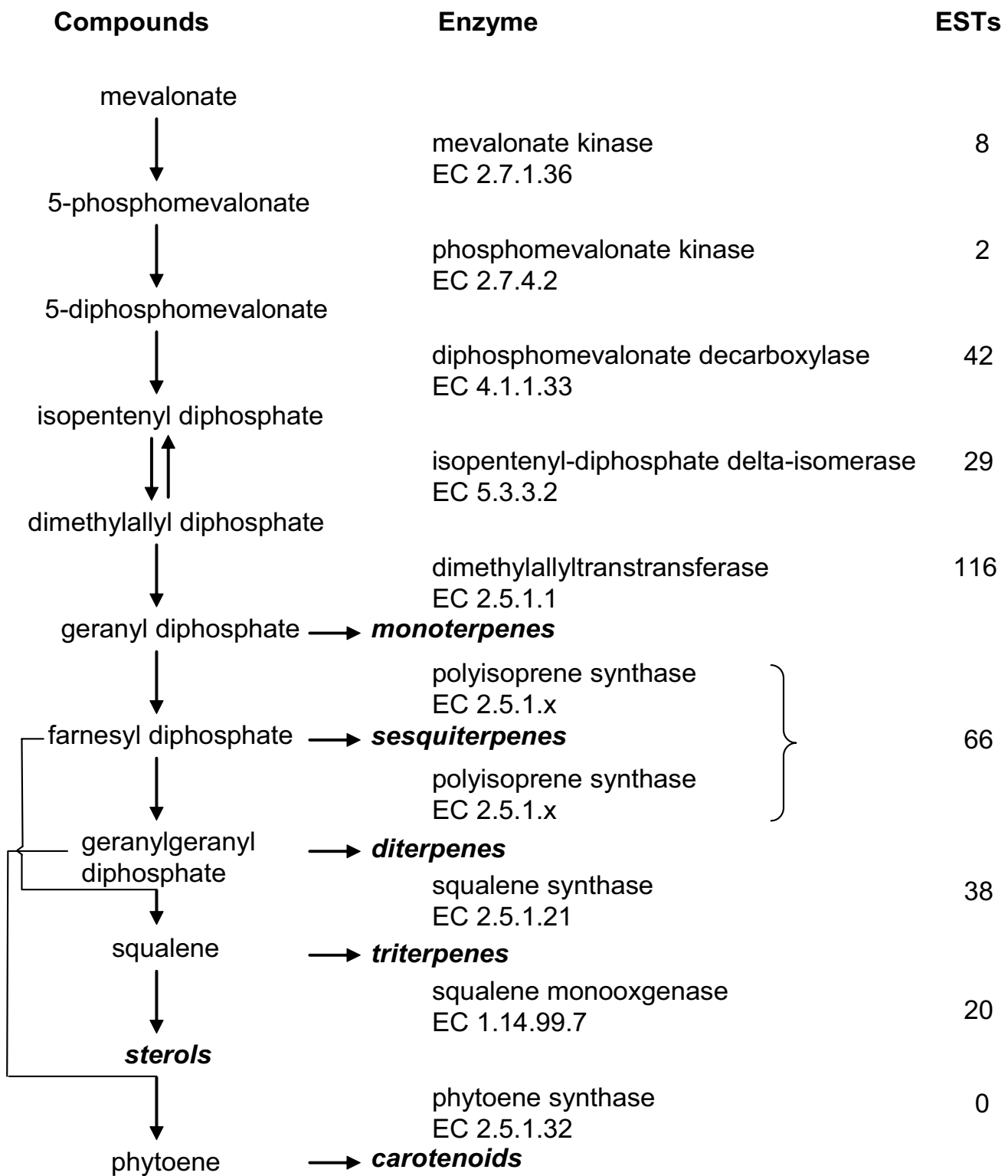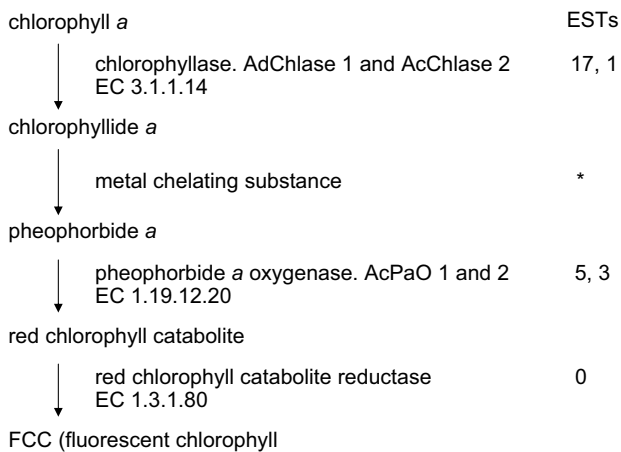phytoene synthase EC 2.5.1.32 — 0

↓

phytoene ⟶ **carotenoids**

Figure 5
Distribution of *Actinidia* ESTs in the terpene biosynthesis pathway. Previously published sequences in GenBank belonging to the terpene biosynthetic pathway from mevalonate [43] were used to identify genes in the *Actinidia* EST database using BLAST searches. ESTs refers to the number of ESTs found in the *Actinidia* EST database for each step of the pathway.

chlorophyll *a*		ESTs
	chlorophyllase. AdChlase 1 and AcChlase 2 EC 3.1.1.14	17, 1
chlorophyllide *a*		
	metal chelating substance	*
pheophorbide *a*		
	pheophorbide *a* oxygenase. AcPaO 1 and 2 EC 1.19.12.20	5, 3
red chlorophyll catabolite		
	red chlorophyll catabolite reductase EC 1.3.1.80	0
FCC (fluorescent chlorophyll		

* No genes known (Hörtensteiner, 2006).

Figure 6
Distribution of *Actinidia* ESTs in the chlorophyll breakdown pathway. Previously published sequences in GenBank belonging to the chlorophyll breakdown pathway of higher plants [46] were used to identify genes in the *Actinidia* EST database using BLAST searches. ESTs refers to the number of ESTs found in the *Actinidia* EST database for each step of the pathway. Where more than one number of ESTs is given, each number represents ESTs that matched a different gene for that step.

way [46]. *Actinidia* ESTs discovered in the pathway of chlorophyll breakdown are shown in Fig. 6. Chlorophyllase is the first enzyme of the catabolic pathway and it removes the phytol chain from chlorophyll *a*. Homologs of this enzyme have been mainly identified in libraries of vegetative tissues (breaking buds, leaves and petals) of *A. deliciosa* and *A. chinensis*. Homologs of pheophorbide *a* oxygenase have also been identified, with eight ESTs representing two genes from fruit libraries of *A. chinensis*. This enzyme is responsible for opening the tetrapyrrolic ring of the chlorophyll molecules, the critical step for the loss of green color. Surprisingly, we did not identify any ESTs of red chlorophyll catabolite reductase, an enzyme associated with pheophorbide *a* oxygenase and required to complete the porphyrin ring opening and the degradation of chlorophyll.

Carotenoid biosynthesis
The fruit of many *Actinidia* species contain a range of carotenoid pigments, including β-carotene and lutein [45]. The yellow-fleshed kiwifruit *A. chinensis* 'Hort16A' develops as a green-fleshed fruit, but on ripening the chlorophyll is degraded exposing the carotenoids in the yellow flesh. ESTs for most steps in the biosynthesis of carotenoids [47-49] are present in the *Actinidia* EST database (Fig.

7), except the first step, phytoene synthase. Ninety eight ESTs for phytoene desaturase were identified, with 94 coming from *A. deliciosa* libraries, and the remainder from *A. chinensis* libraries. Of these 98 ESTs, 90 came from a single *A. deliciosa* fruit library with eight from bud and meristem libraries. Of the 56 ESTs found for lycopene β-cyclase, 23 were from petal tissues and 32 from fruit libraries. These ESTs were largely found in libraries from *A. deliciosa* (31) and *A. chinensis* (24), with only one EST from *A. eriantha*. In contrast, only a single EST was found for lycopene ε-cyclase, which is the enzyme required with lycopene β-cyclase to convert lycopene to α-carotene [50]. There were five ESTs of β-carotene hydroxylase found in petal (3), leaf (1) and dormant bud (1) libraries, but only a single EST was present for ε-carotene hydroxylase. Both of these enzymes are required for lutein synthesis [51].

Flavonoid and anthocyanin biosynthesis
Flavonoids are polyphenolic secondary metabolites synthesized through the phenylpropanoid pathway [52]. Anthocyanins are a subgroup of the flavonoids, which appear red to blue, depending on the pH, due to additions to their phenolic rings. In red fruit of *A. chinensis* and *A. deliciosa*, the predominant anthocyanins are cyanidin-based and the preferred glycosylation is in the 3' position [53]. While many of the kiwifruit libraries came from tissue with low levels of anthocyanins, several libraries were derived from tissue containing anthocyanins, including red fruits, brown skins and pink petals, leaves and buds (Additional file 2a).

The anthocyanin biosynthetic pathway was analyzed by BLAST searching for ESTs representing different enzymes in the pathway (Fig. 8). Chalcone synthase (CHS) was a highly represented gene family in the EST database with ~300 ESTs. There appeared to be three distinct CHS NRs, possibly representing three genes, that are found in mostly young *A. eriantha* fruit libraries, as well as in bud and petal libraries. There was only one apparent chalcone isomerase (CHI) NR and over half the ESTs for this gene were found in shoot and bud libraries. Two distinct flavanol 3-hydroxylase (F3H) NRs were found, with ESTs predominantly found in bud and cell culture libraries. Only one flavonoid 3',5'-hydroxylase (F3'5'H) NR was found, from *A. eriantha* fruit, while ESTs for two distinct flavonoid 3'-hydroxylase (F3'H) NRs were predominantly found in shoot and bud libraries. Of the four potential dihydroflavonol reductase (DFR) NRs, the most highly represented was found in bud and young fruit. Three NRs, perhaps representing three gene families, of leucoanthocyanidin dioxygenase (LDOX) were found in the EST database, with most ESTs in bud libraries. The ESTs with highest sequence similarity to UDP-glucose flavonoid 3-O-glucosyltransferase (UFGT) are also predominantly found in developing buds of *A. deliciosa*.
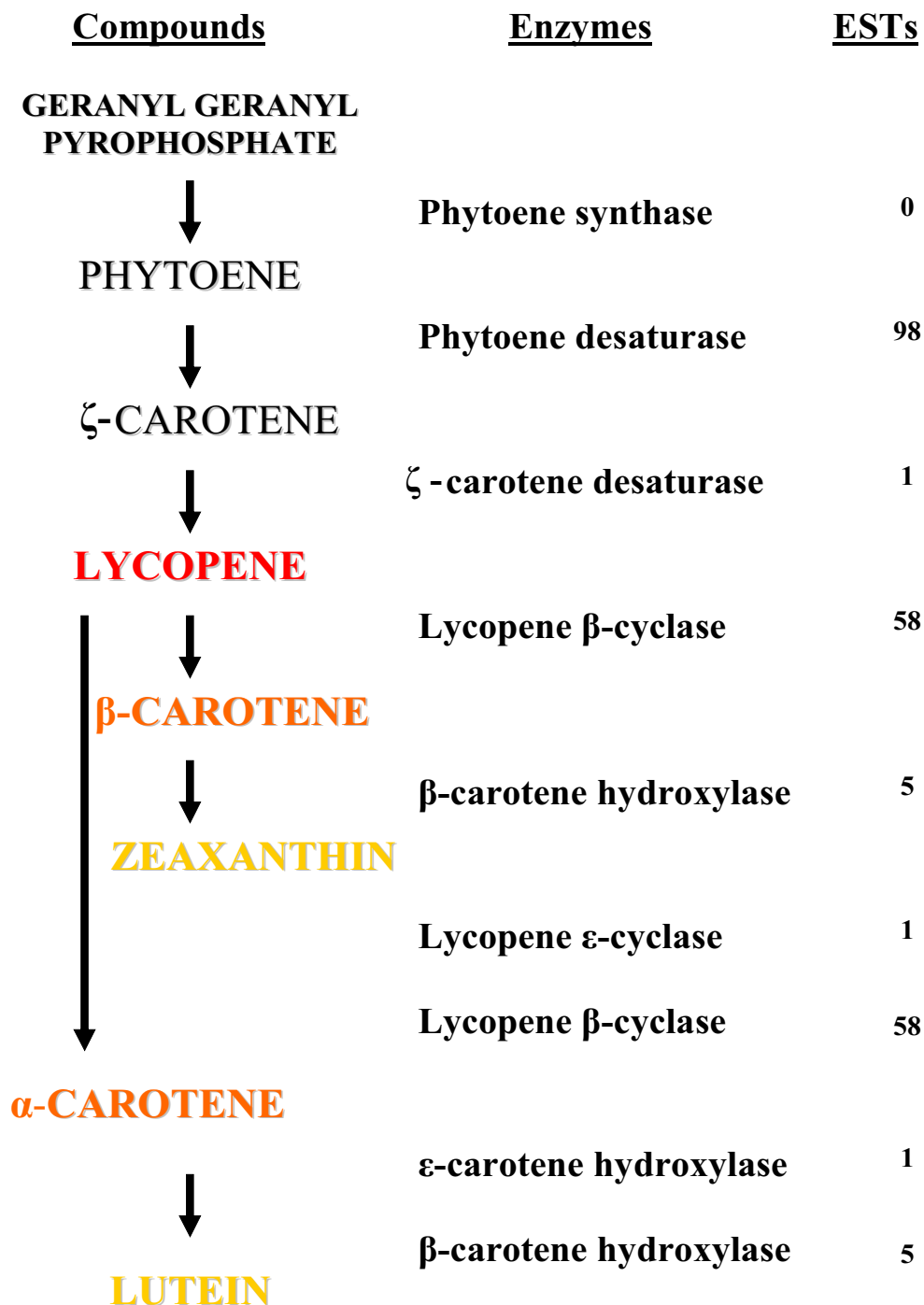
Carotenoid Biosynthesis Pathway

Compounds	Enzymes	ESTs
GERANYL GERANYL PYROPHOSPHATE		
↓	**Phytoene synthase**	**0**
PHYTOENE		
↓	**Phytoene desaturase**	**98**
ζ-CAROTENE		
↓	ζ -carotene desaturase	**1**
LYCOPENE		
↓	**Lycopene β-cyclase**	**58**
β-CAROTENE		
↓	**β-carotene hydroxylase**	**5**
ZEAXANTHIN		
	Lycopene ε-cyclase	**1**
	Lycopene β-cyclase	**58**
α-CAROTENE		
↓	**ε-carotene hydroxylase**	**1**
	β-carotene hydroxylase	**5**
LUTEIN		

Figure 7

Distribution of *Actinidia* ESTs in the carotenoid biosynthetic pathway. Previously published sequences in GenBank belonging to the carotenoid biosynthetic pathway [47] were used to identify genes in the *Actinidia* EST database using BLAST searches. ESTs refers to the number of ESTs found in the *Actinidia* EST database for each step of the pathway.
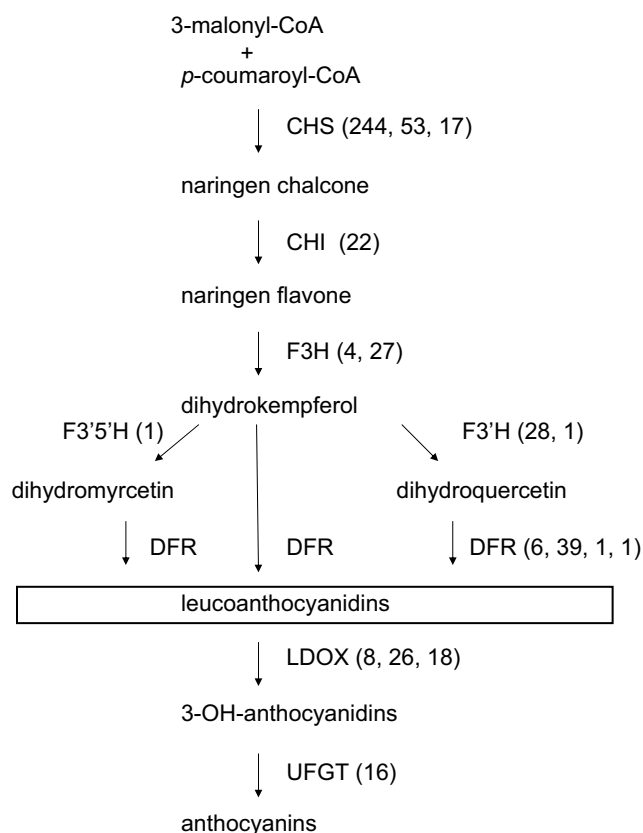
3-malonyl-CoA
+
p-coumaroyl-CoA

↓ CHS (244, 53, 17)

naringen chalcone

↓ CHI (22)

naringen flavone

↓ F3H (4, 27)

dihydrokempferol

F3'5'H (1) ↙   ↘ F3'H (28, 1)

dihydromyrcetin   dihydroquercetin

↓ DFR   ↓ DFR   ↓ DFR (6, 39, 1, 1)

leucoanthocyanidins

↓ LDOX (8, 26, 18)

3-OH-anthocyanidins

↓ UFGT (16)

anthocyanins

Figure 8
Distribution of *Actinidia* ESTs in the flavonoid biosynthetic pathway. Previously published sequences in GenBank belonging to the flavonoid biosynthetic pathway [49] were used to identify genes in the *Actinidia* EST database using BLAST searches. Enzyme names are followed in brackets by the number of ESTs in different families of genes encoding these enzymes. See text for abbreviations.

Many of the ESTs for anthocyanin biosynthesis were sequenced from young fruit libraries (particularly *A. eriantha*) and in *A. deliciosa* buds. Developing buds from *A. deliciosa* appear to be a darker red color and therefore presumably accumulate more anthocyanins than buds from *A. chinensis* and *A. eriantha*, so it is not surprising to find ESTs for these later biosynthetic steps in bud tissue from *A. deliciosa*. However, ESTs for genes encoding enzymes from the early steps in the flavonoid pathway (e.g. CHS and F3H) are found in unpigmented *A. eriantha* fruit suggesting these tissues may accumulate a different class of flavonoids.

Most of the enzymes for anthocyanin biosynthesis are well represented in the EST database except for F3'5'H. This is not surprising given the only anthocyanins identified in *A. deliciosa* and *A. chinensis* are glycosylated in the 3' position. There are, however, species within the genus

Actinidia, such as *A. melanandra*, that accumulate high levels of delphinidin, a different aglycone that would require F3'5'H activity. No ESTs from these species were sequenced.

Genes related to healthful components
Ascorbic acid-related genes
Kiwifruit are well known to be high in ascorbic acid (vitamin C), with values in commercial varieties ranging from 80 to 120 mg per 100 g fresh weight of fruit [54], and with *A. eriantha* having up to 10 times as much again. Ascorbic acid is synthesized in plants from glucose, through three potential pathways: the L-galactose pathway, the galacturonate pathway and through *myo*-inositol [55]. *Myo*-inositol is a major sugar alcohol found in kiwifruit (Additional file 7; [56,57]). The only pathway that is well established is the L-galactose pathway for which all the enzymes have now been identified in plants [55,58]. Many of the genes are known for the other two potential pathways, but several steps are still unknown. The *Actinidia* EST database was searched for ESTs similar to genes known to encode steps in these pathways. The numbers of ESTs homologous to various identified genes in the three pathways of ascorbic acid biosynthesis are shown in Fig. 9. Surprisingly for these high vitamin C plants, while the number of ESTs found in early steps in all the pathways is relatively large, the number of ESTs found in the two last steps is much lower, with no examples of ESTs encoding the last step of two of the pathways, that catalyzed by L-galactono lactone dehydrogenase.

Quinic acid metabolism
Fruit from *Actinidia* species have a relatively high total acid content (1–3% w/w) of which 40–60% can be quinic acid, 40–60% citric acid and 10% malic acid (Additional file 7). Quinic acid comprises an even higher proportion of the total acids during early fruit development [59], and over 25-fold variation in quinic acid content can be found in fruit of a single *A. chinensis* cross [60]. Kiwifruit are unusual in having such a high content of quinic acid as a fresh fruit, although levels of quinic acid in cranberry juice can be around 1% [61] and peaches can have up to 0.2% quinic acid [62].

Quinic acid is a key intermediate in lignin biosynthesis, folic acid metabolism, aromatic acid synthesis, anthranilate biosynthesis, and purine metabolism. The molecular and enzymatic control of quinic acid storage and metabolism may affect all these pathways. However, little is known about quinic acid metabolism in *Actinidia* species. ESTs for all the enzymes in the quinic acid biosynthetic pathway to shikimate (Fig. 10), except for a quinate dehydrogenase, have been identified in the *Actinidia* EST database.
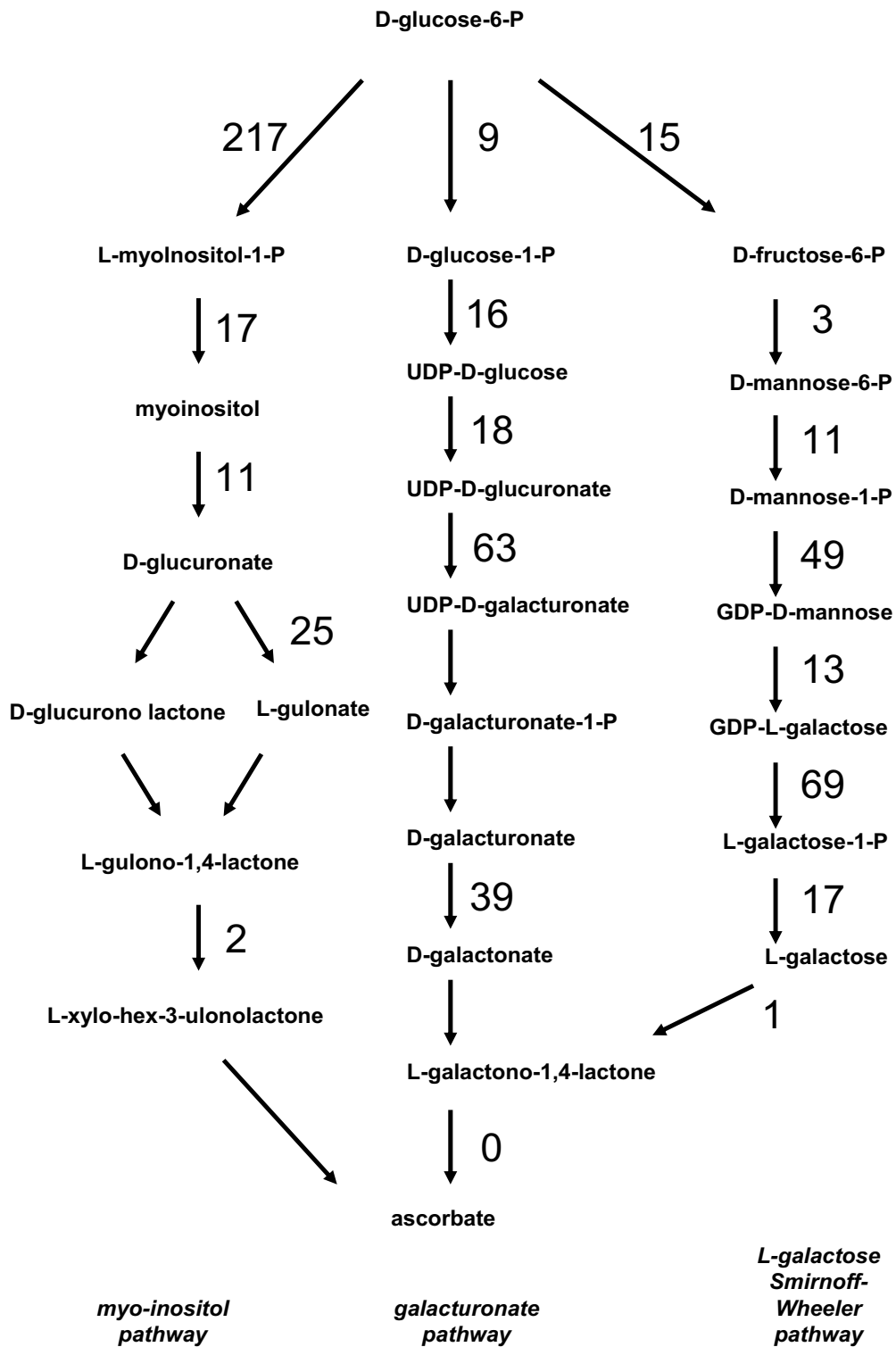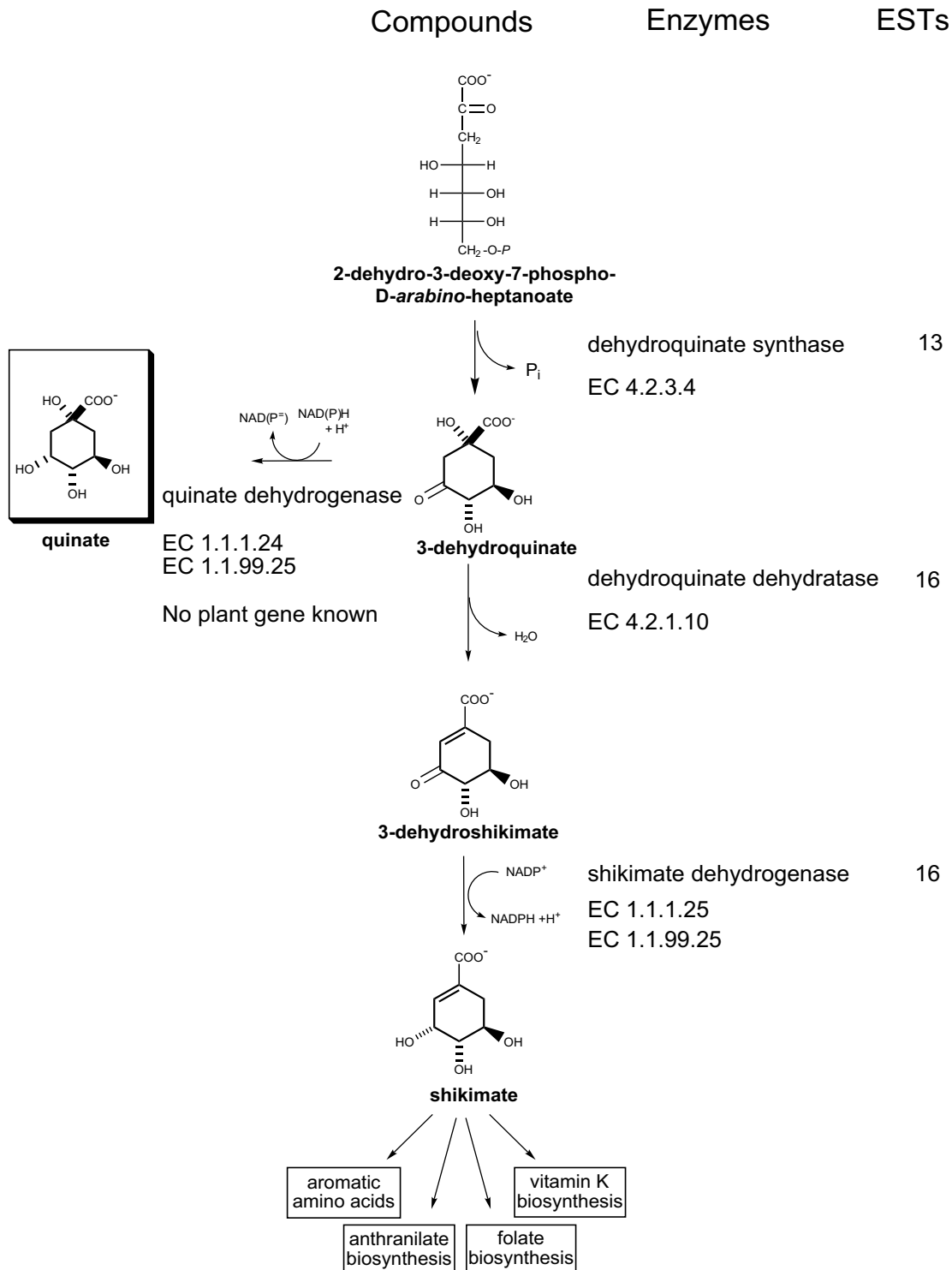
D-glucose-6-P

217 / 9 / 15

L-myoInositol-1-P **D-glucose-1-P** **D-fructose-6-P**

17 16 3

myoinositol **UDP-D-glucose** **D-mannose-6-P**

11 18 11

D-glucuronate **UDP-D-glucuronate** **D-mannose-1-P**

25 63 49

D-glucurono lactone **L-gulonate** **UDP-D-galacturonate** **GDP-D-mannose**

13

D-galacturonate-1-P **GDP-L-galactose**

69

L-gulono-1,4-lactone **D-galacturonate** **L-galactose-1-P**

2 39 17

L-xylo-hex-3-ulonolactone **D-galactonate** **L-galactose**

1

L-galactono-1,4-lactone

0

ascorbate

myo-inositol pathway *galacturonate pathway* *L-galactose Smirnoff-Wheeler pathway*

Figure 9
Distribution of *Actinidia* ESTs in three ascorbic acid biosynthetic pathways. Previously published sequences in GenBank belonging to the three known ascorbic acid biosynthetic pathways in plants [55] were used to identify genes in the *Actinidia* EST database using BLAST searches. The number of ESTs found in the Actinidia EST database for each step of each pathway is shown. Steps where no ESTs are noted have not had genes identified.

Compounds   Enzymes   ESTs

**2-dehydro-3-deoxy-7-phospho-
D-*arabino*-heptanoate**

dehydroquinate synthase   13

EC 4.2.3.4

3-dehydroquinate

quinate dehydrogenase

EC 1.1.1.24
EC 1.1.99.25

No plant gene known

quinate

dehydroquinate dehydratase   16

EC 4.2.1.10

3-dehydroshikimate

shikimate dehydrogenase   16

EC 1.1.1.25
EC 1.1.99.25

shikimate

aromatic
amino acids

anthranilate
biosynthesis

folate
biosynthesis

vitamin K
biosynthesis

Figure 10
Distribution of *Actinidia* ESTs in the quinate biosynthetic pathway. Previously published sequences in GenBank belonging to the quinate biosynthetic pathway [98] were used to identify genes in the *Actinidia* EST database. The number of ESTs found in the *Actinidia* EST database for each step of each pathways is shown. Dehydroquinate dehydratase and shikimate dehydrogenase are two activities of a single bifunctional enzyme encoded by one gene [99].

Of the 13 ESTs identified for dehydroquinate synthase, eight were found in fruit or petal libraries, the rest were sequenced from bud or leaf libraries. Of the 16 ESTs identified for the bifunctional dehydroquinate dehydratase/shikimate dehydrogenase, 11 were from fruit or petal libraries and five from bud or leaf libraries. These ESTs were found mainly in libraries made from *A. deliciosa* and *A. chinensis* tissues and to a lesser extent in *A. arguta* and *A. eriantha* libraries, reflecting the distribution of ESTs among the species. An enzyme has been reported with predominantly quinate dehydrogenase activity [63], but the gene for this enzyme has not been cloned.

Allergens
Food allergies are estimated to affect ~6% of young children and ~3% of adults [64]. Kiwifruit have been recognized as a potentially allergenic fruit for over 20 years and reported allergies to kiwifruit are increasing [65]. Most allergy symptoms to kiwifruit are quite mild, but severe reactions have been reported, particularly in young children [66]. Allergies have been recorded in all three commercial species of *Actinidia* [67].

The *Actinidia* EST database contains sequences with homology to many known plant allergen proteins [68]. Potential allergens include the 2S albumin proteins, lipid transfer proteins, thaumatin-like protein (TLP), α-amylase/trypsin inhibitors, latex allergens, plant chitinases, profilins, cystatins, Bet v 1 homologous proteins and plant seed globulin allergens. However, only a small number of potential kiwifruit allergens have been directly confirmed by immunological testing, including the cysteine protease actinidin (Act d 1; [69]), an unidentified 43 kDa protein (Act d 2; [70], a TLP [70] and kiwellin [25].

ESTs for actinidin occurred throughout the 37 libraries from seven species that make up the database, and these clustered into acidic and basic forms of 10 distinct genes/alleles [22]. No matches were found by BLAST searching the EST database for the 43 kDa protein Act d 2. Kiwellin ESTs were found in libraries from most *Actinidia* species but were particularly abundant in *A. eriantha* ripe fruit skin, accounting for ~40% of ESTs in that library. The EST database contained 24 NRs with homology to TLPs from other plants (e.g. Pru a 2 from cherry and thaumatin 1 from *Thaumatococcus daniellii*), and all but one EST contained the 16 conserved cysteine residues characteristic of this protein class [71]. The various NRs could be aligned and clustered into 4 or 5 groups (data not shown), with two NRs accounting for the bulk of the ESTs (46 and 29%).

Globulin proteins (11S and 7S) have been reported from a wide range of seeds, including cereals [72,73] and leg-

umes [74]. Globulin-like proteins are also found in *A. thaliana* [75] and a gene family of 10 globulin-like genes are present in *Arabidopsis* gene databases (The *Arabidopsis* Information Resource (TAIR), http://arabidopsis.org/). 11S globulins are composed of sets of polymorphic subunits derived from a multigene family. They exist as hexamers with molecular mass between 300 and 400 kDa, or as trimers. 11S globulin subunits consist of two polypeptide chains linked by at least one S-S bridge between cysteine residues at highly conserved positions in the acidic alpha-chain and basic beta-chain. Both chains are post-translationally generated from a common precursor protein, which represents the product of one member of the multigene family [76]. 11S globulins have been shown to be related to the 7S globulins through sequence comparisons [77].

A BLAST search of the *Actinidia* EST sequence database with a range of previously identified globulin protein sequences identified two ESTs with a predicted protein sequence matching 11S globulin, both from a library made from *A. chinensis* whole fruit, which included seeds. Other related sequences from *Actinidia* fruit libraries belonged to the 7S vicilins-like family (38 ESTs), which were found in only fruit-based libraries from *A. deliciosa*, *A. chinensis* and *A. arguta*. These libraries would have all included seeds. A third grouping of sequences was identified that we have named other globulins (OG) (201 ESTs). The OG group appears to be more closely related to the 7S globulins than to the 11S (Fig 11). All three groupings are supported by the clustering of the *Arabidopsis* globulin genes as well as globulin genes from other species. While most OG ESTs were found in fruit libraries (81%), which included seeds, some members of the OG family were found in fruit skin, buds and petal libraries, suggesting a role other than as seed storage proteins. These OG ESTs were found in libraries derived from *A. deliciosa* (100), *A. chinensis* (42), *A. arguta* (39), and *A. eriantha* (16).

Genes related to fruit softening
Some kiwifruit cultivars exhibit outstanding storage characteristics e.g. *A. deliciosa* 'Hayward' fruit can be stored for 4–6 months at 0°C. Other *Actinidia* species exhibit a range of ripening and softening behaviors [78]. *A. eriantha* develops a peelable skin as the fruit ripen, *A. arguta* ripens in less than 10 days at 20°C versus 20–25 days for *A. deliciosa* and *A. chinensis* [79], whilst some small-fruited *Actinidia* genotypes tend to remain firm even towards the end of the ripening process [78].

The main chemical changes occurring in the cell wall during kiwifruit softening are pectin solubilization and degradation, reduction in the molecular weight of xyloglucan, and galactose loss from pectin side chains.
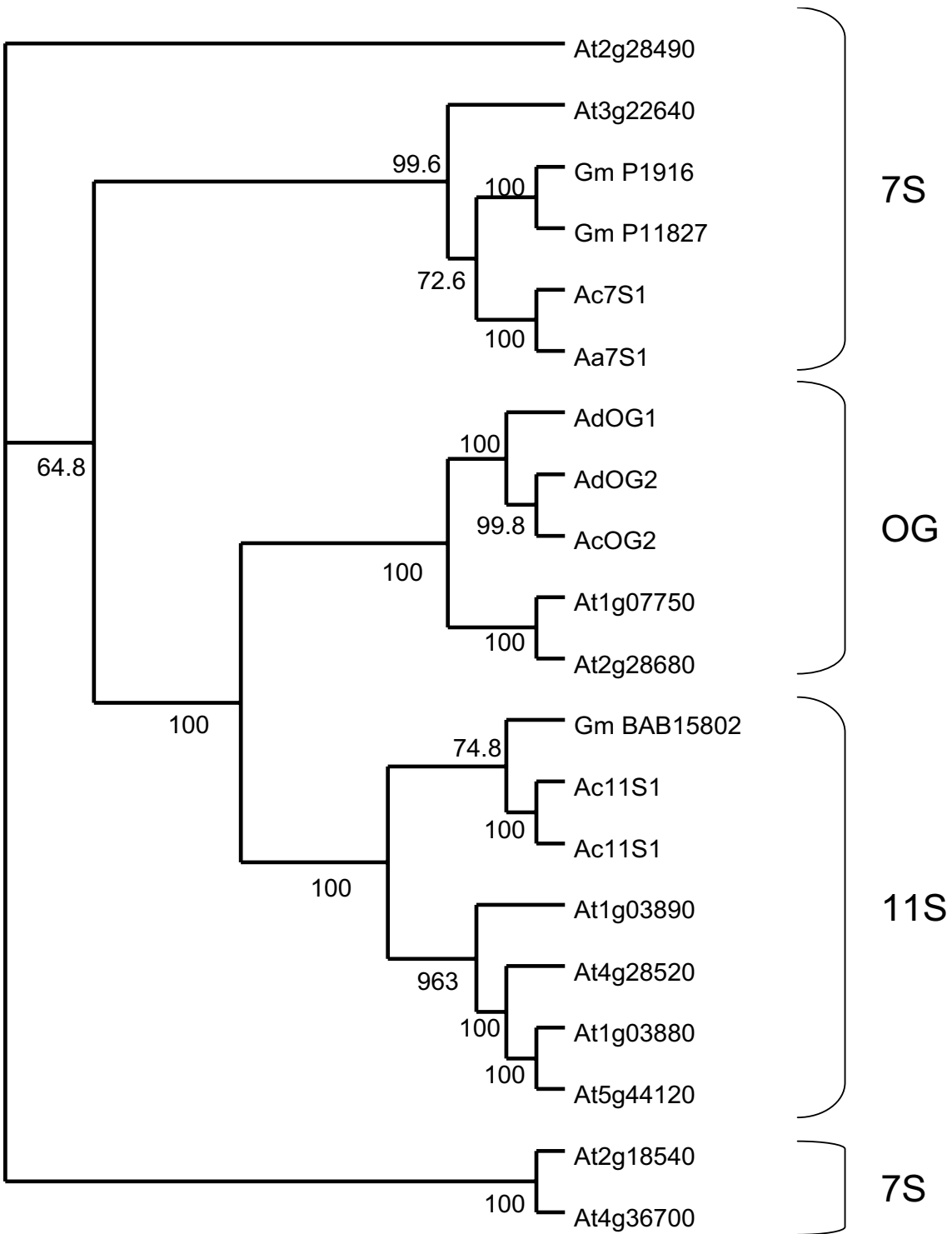
Figure 11
Phylogenetic tree of plant globulins of known function and *Actinidia* members of the globulin family. Previously published plant globulin sequences from GenBank were used to identify genes from the *Actinidia* EST database using BLAST searches. Abbreviations for species are as follows: Aa, *Actinidia arguta*; Ac, *A. chinensis*; Ad, *A. deliciosa*; At, *Arabidopsis thaliana*; Gm is *Glycine max* (soybean) followed by the Genbank accession number. Percentage bootstrap values (1000 bootstrap replicates) for groupings are given by each branch.
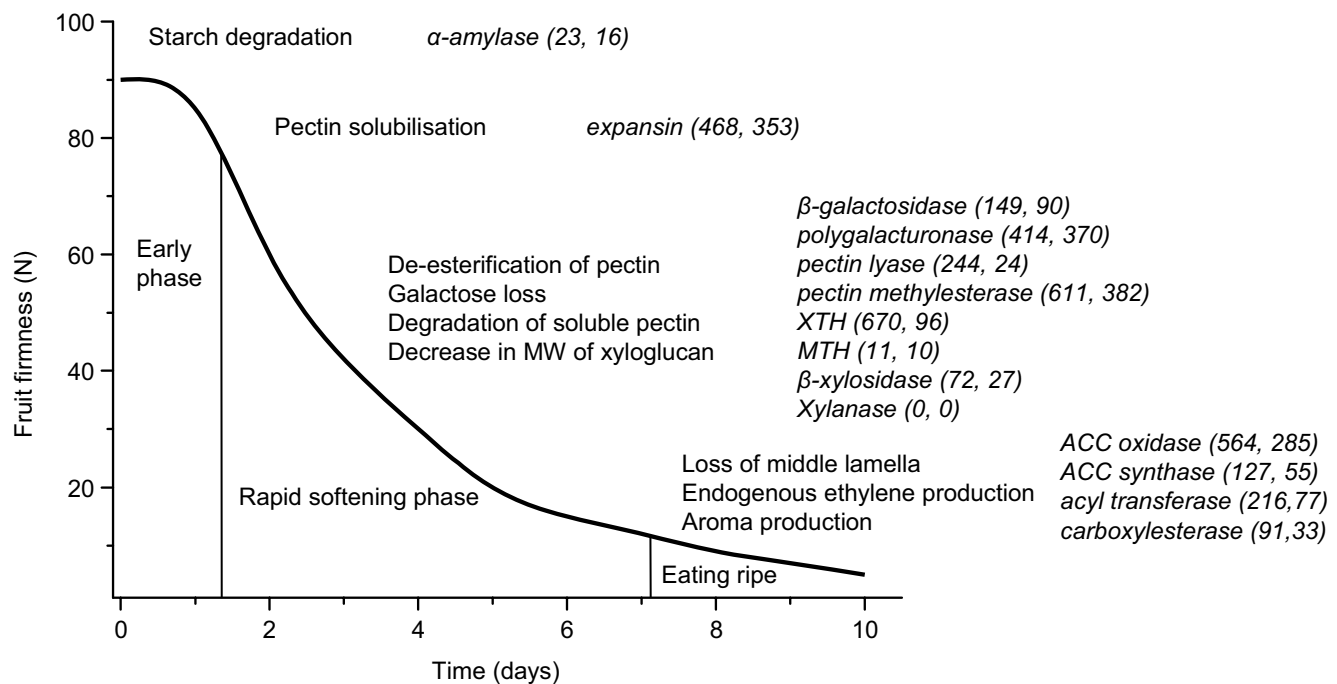
Figure 12
Distribution of *Actinidia* ESTs involved in postharvest kiwifruit softening. Schematic representation of postharvest kiwifruit softening in relation to the timing of key events in the softening process of ethylene-treated 'Hayward' fruit. Diagram modified from [80]). Previously published sequences in GenBank involved in postharvest softening processes were used to identify genes in the *Actinidia* EST database using BLAST searches. The first number in each set of brackets represents the total number of ESTs found in the *Actinidia* database and the second number is the number of ESTs found in fruit libraries.

Microscopically, the cell wall shows extensive swelling, until at the end of the softening process dissolution of middle lamellae can be observed [80]. These processes eventually lead to disintegration of the cell wall. Although similar chemical changes occur in other fruit [81], these changes occur concurrently with ethylene production and the respiratory climacteric. In kiwifruit, however, most of the softening process is well separated temporally from the climacteric and ethylene production (shown schematically in Fig. 12).

The genes for many enzymes involved in the key chemical changes in the cell wall during ripening are represented in the *Actinidia* EST database. Three enzymes involved in pectin degradation (pectin lyase, pectin methylesterase and polygalacturonase) are particularly abundant with 244, 611 and 414 ESTs, respectively. Genes encoding expansin, a protein implicated in pectin solubilization early in the kiwifruit softening process [80], are also highly abundant (468 ESTs). Galactose loss from fruit cell walls during the softening process has been attributed to the action of β-galactosidases [82]. A β-galactosidase from

ripe kiwifruit has been purified. However, the activity of the enzyme *in vitro* against synthetic or purified native substrates from kiwifruit was far too low to account for the amount of galactose loss observed during softening [83]. Ninety β-galactosidase ESTs were observed in fruit libraries in the EST database and are good candidates for genetic manipulation and to study the role of this enzyme. During softening, xyloglucan is also hydrolyzed. As xyloglucan and cellulose create the major load-bearing network of the cell wall, any xyloglucan degradation is thought to weaken the cell wall, resulting in softening of the fruit. In kiwifruit, hydrolysis of xyloglucan occurs mainly during the rapid softening phase. Endoglucanases and xyloglucan transglucosylase/hydrolase (XTH) enzymes are implicated in this process. XTHs catalyze both hydrolytic and transglucosylation reactions, and an XTH enzyme capable of carrying out both reactions has been isolated from kiwifruit [84]. This particular XTH, however, is expressed late in kiwifruit ripening, when xyloglucan degradation has already come to an end. The 96 XTH ESTs found in fruit libraries will allow the identification of genes involved in hydrolysis of xyloglucan during the rapid sof-

tening phase. ESTs for enzymes that act on minor kiwifruit cell wall components such as galactoglucomannans or glucuronoarabinoxylans [80] were less abundant in the EST database (e.g. mannan transglycosylase/hydrolase (MTH), 10 ESTs found in fruit).

Discussion

We report on a significant resource of over 130,000 ESTs derived from a range of *Actinidia* species (Table 1). We targeted tissues and developmental stages in order to sample genes involved in physiological and biochemical processes including fruit ripening, flavor development, control of color and the synthesis of chemicals with health-related attributes. For this reason, the two most widely cultivated species of kiwifruit, *A. chinensis* and *A. deliciosa*, are well represented with together over 100,000 ESTs (Table 1). In addition, fruit and bud libraries are also well represented, with over 38,000 and 50,000 ESTs, respectively. *A. chinensis* and *A. deliciosa* are so closely related, as is *A. setosa* (Li) C.F. Liang et A.R.Ferguson, that they are variously treated as being distinct species or as varieties of the one species. The other two main species studied, *A. arguta* and *A. eriantha* Benth., also have commercial potential but are more distantly related [2].

The genus *Actinidia* is unusual in how much inter-taxal and intra-taxal variation in ploidy and in the wild, there is a structured reticulate pattern of diploids, tetraploids, hexaploids, and octoploids in diminishing frequency, associated, in at least some taxa, with geographic separation of ploidy races. *A. deliciosa* is hexaploid, *A. setosa* is diploid, and there are diploid and tetraploid races of *A. chinensis*, the tetraploids apparently coming from a restricted part of the natural distribution of the species. Most evidence suggests that diploid *A. chinensis* was a progenitor of tetraploid *A. chinensis* and hexaploid *A. deliciosa* but it is not clear whether genomes from other species have contributed. The basic chromosome number (n = 29) is high and it seems increasingly likely that diploid *A. chinensis* is itself a rediploidized palaeopolyploid [2].

As is common in EST sequencing projects (e.g., [6,8,10]), there is a high degree of redundancy in the ESTs, with clustering reducing the number of unique sequences from over 132,000 to 41,858 NRs (18,070 TCs, 23,788 singletons). We would expect this number of NRs to be an overestimate of the number of genes in *Actinidia*, especially given that the database contains sequences from multiple species of *Actinidia*. Using the same correction used in the apple EST paper [10], we expect an *Actinidia* genome to have around 27,000 genes.

On average 20% (± 2% standard error) of the sequences from each library with over 1000 ESTs were singletons

suggesting a high degree of novelty in these libraries. On average 28% (± 4%) of sequences did not have a homolog in the various public databases based on BLAST searches with an E value > 1.0e^{-10}. An average of 16% (± 3%) of ESTs were identified as 3' UTR candidates based on the presence of a poly(A) tail within 40 bp of the start (taking into account reverse sequences). These 3' sequences would not be expected to be identified by BLAST searches and so would affect the novelty of a library. Less than 12% of NRs did not have BLAST matches (E > 10) in the *Arabidopsis* proteome, Uniref, NCBI ref or SwissProt databases.

There was only a small degree of overlap in NRs between libraries. Libraries from different species and different tissues showed a 5 to 9% overlap in NRs, libraries from different species but the same tissue showed an 8 to 10% overlap and libraries from the same species but different tissue showed a 7 to 13% commonality in NRs. These comparisons were made over five large libraries with more than 9,000 EST members each and an average of 2.1 ESTs per NR. These results suggest that there were more NRs in common between libraries made from the same tissue or from the same species, but this tendency was not particularly marked.

Detecting SNPs using an automatically assembled EST database is a cost effective way to discover new DNA polymorphisms and develop novel markers, although it can be a challenging task, especially in polyploid *Actinidia* species. A significant proportion of the sequence variants predicted from overlapping ESTs within an NR will correspond to "real" SNPs, which means the sequence differences found are allelic variants of a given locus and not sequencing errors or differences between paralogs, homoeologs or orthologs. Homoeolog SNPs could be particularly common in the polyploid accessions of species such as *A. deliciosa* and *A. arguta* that make up a large proportion of this database, but are also possible in diploids as a result of conserved gene pairs of paleopolyploid origin. Allelic SNPs can be used directly and converted into molecular markers for genetic mapping, population genetics and linkage disequilibrium studies or for marker-assisted selection. A SNP marker for determining the sex of kiwifruit seedlings [85] has already been successfully utilized. Since the database contains sequence data from multiple species, and ~40% of TCs are made up of more than one species, several SNPs were detected in the *Actinidia* EST database corresponding to sequence between orthologous loci from different *Actinidia* species. Hence, they cannot fully be considered as allelic SNPs, but more as species-specific variations. However, as kiwifruit breeding programs often use controlled crosses between different species, the interspecific SNPs will segregate in the progeny and be useful as markers.

The incidence of SSRs in NRs was higher in *Actinidia* (30%) than in apple (20%), and the frequency of di-nucleotide and tri-nucleotide SSRs differed between these two species. This increase was evident in all of the sequence classes but greatest in AG and AC (double the incidence among apple NRs). Even though the *Actinidia* genome EST resource represents several species, while apple came mainly from one species, this would not explain these differences. Perhaps the longer period of domesticity in apple, based on a narrow genetic basis compared to kiwifruit, may explain the difference. Alternatively it may reflect that a greater proportion of homoeologs have grouped into TCs in the polyploid kiwifruit data than in the apple dataset.

Overall the codon usage of the three *Actinidia* species shares many similarities with that of other dicotyledons represented in the codon usage database [86]. Comparisons with *Arabidopsis* codon usage showed that *A. deliciosa* and *A. eriantha* differ markedly for 15 and 17 amino acids, respectively, whereas *A. chinensis* differs in its preference for eight particular amino acids. Further comparisons with apple, grape, pear, peach, loblolly pine, tomato, citrus, potato and tobacco showed that the codon preference of the *Actinidia* species is most similar to that of apple [10]. *A. deliciosa* differs from apple only in its codon preference for aspartate, glycine, isoleucine and leucine. *A. eriantha* also differs from apple for these four amino acids and also serine. The codon preference for *A. chinensis* and apple also differ for only four amino acids, these being asparagine, glutamine, threonine and valine. CpG suppression is also evident in *Actinidia* species with an XCG/XCC ratio of between 0.68 and 0.71 for the three species evaluated. This modest level of suppression of the CpG di-nucleotides is similar to that of apple (0.64) and differs markedly from that of *Arabidopsis* which shows nearly no suppression (0.92) and from the high level found in grape (0.35). This may well reflect different levels of methylation in the coding sequences used by different species of plants.

Mapman was used to assign function to the *Actinidia* NRs and thus to their constitutive ESTs. Only 32% of the ESTs did not have an *Arabidopsis* homolog at E < 1.0e^{-10}. In general, the functional distribution of NRs was very similar to the functional distribution of *Arabidopsis* proteins (Table 3) suggesting that the sampling of *Actinidia* ESTs well represented the major functional classes of plant genes. This is surprising given the biased selection of libraries with virtually no root ESTs sequenced. However, the high number of bud meristem libraries meant that genes expressed in metabolically active dividing tissue were sampled.

Fruit of the *Actinidia* genus show several characteristics that distinguish them from other fruit species. These include flesh color (green is the most common, but yellow, orange and red fruit also occur in the genus), chemical composition including high vitamin C and quinic acid contents, and a novel aroma composition (Additional file 7), characterized by the abundant esters. In addition, kiwifruit has been identified as a fruit with a potential to cause allergenicity among consumers, although this is a problem common to many other fruit. For this reason, we analyzed the *Actinidia* EST database to identify genes involved in these pathways and products. These analyses showed the depth and usefulness of the database for selecting candidate genes for most steps in the selected pathways. The other useful characteristic of the *Actinidia* EST database is the wide range of genetic and phenotypic diversity sampled across the *Actinidia* genus (Fig. 1) and the value of using this diversity to discover novel traits through functional genomics and through mapping and positional cloning approaches.

Conclusion

This paper describes an EST resource in the *Actinidia* genus and discusses many of the properties of this collection. However, there is still a tremendous challenge in understanding the molecular basis of the genetic diversity of this genus, and we expect putting this EST resource into the public domain will enhance future understanding of the genetic basis of the many divergent traits in this fruit.

Methods
Plant material
Tissues were collected from *Actinidia* species growing in New Zealand at HortResearch research orchards in Auckland, Bay of Plenty (Te Puke) and Northland (Kerikeri) from 1999 to 2003. Table 1 and Additional file 2a provide details of tissues, species and treatments.

Library construction and EST sequencing
Total RNA was extracted from *Actinidia* tissues by established methods [87,88]. Messenger RNA was isolated from total RNA by passage through oligo(dT)-cellulose columns (GE Healthcare, USA), and cDNA cloned into either phage (ZapcDNA Synthesis Kit and Zap-cDNA Gigapack III Gold Cloning Kit; Stratagene, USA) or plasmid-based libraries (SuperScript System for cDNA Synthesis and Cloning; Invitrogen, USA). In some libraries, subtractive or other enrichment techniques were used (Additional file 2b).

Plasmids from the phage cDNA libraries were mass excised, according to the manufacturer's recommendations (Stratagene). Plasmid extractions were then under-

taken on individual bacterial colonies of either the phage-derived or the plasmid-derived cDNA libraries and the corresponding cDNA inserts sequenced predominantly from the 5' end. Big Dye Terminator sequencing reactions were resolved on ABI377, ABI3100, or ABI3700 sequencers, according to the manufacturer's instructions (Applied Biosystems, USA). For determination of the complete sequence of cDNA clones, M13R and M13F or T3 and T7 primers were used for 5' and 3' end sequencing. EST-specific primers were used to determine the complete sequence of cDNA clones. In situations where EST clones had long poly(A) tails (generally 40 nucleotides) and, therefore, failed to yield good quality sequence with standard sequencing primers, an anchored T24VN primer was used. Resulting sequences were edited manually and assembled using Sequencher software, version 4.0.5 (GeneCodes). Sequencing progress for each cDNA library was assessed manually for clone length, redundancy and sequence quality.

Bioinformatics

EST sequences were automatically trimmed of vector, adapter, and low quality sequence regions, and uploaded to a relational database. Automatic annotation was performed using the HortResearch BioView sequence annotation pipeline (BioView – an enterprise bioinformatics system for automated analysis and annotation of non-genomic DNA sequence (Crowhurst R, Davy M, Deng C, unpublished)) that utilizes a relational database (MySQL; http://www.mysql.com). The EST clustering phase was performed using The Institute for Genomic Research (TIGR) gene indices clustering tools http://www.tigr.org/tdb/tgi/software. The representation of protein families, domains, and functional sites within the *Actinidia* NRs was determined using InterProScan [89]. The proteome for *Arabidopsis* (*Arabidopsis thaliana*) was obtained from TAIR [90], and comparisons to proteins from *Arabidopsis* using BLASTx were used to identify *Actinidia* NRs with similarity to *Arabidopsis* proteins.

Detection of SSRs was undertaken using a PERL program within BioView that identified tandem repetition of sequence words in target sequences. SSRs were characterized by repeat type (di-, tri-, or tetra-nucleotide repeat units), repeat length, and position. Only repeats longer than 10 bp were included in the analyses. When reporting the frequency of repeat classes, different di- and tri-nucleotide sequences were combined by type; for example, AG repeats also encompassed repeats identified as GA and their complementary sequences CT or TC repeats. Prediction of SNPs and insertion/deletions and sequencing errors was performed using PERL scripts within BioView that parsed the output of contig sequences generated by the CAP3 DNA sequence assembly program [91].

Codon usage tables were derived from cDNA sequences encoding predicted full-length proteins. Clones were predicted to be full length only if they started with an ATG codon and terminated with a stop codon at positions equivalent to those of other plant genes. Codon usage was calculated from sequences using the CUSP program implemented within EMBOSS [92]. For functional analysis using Mapman [27], each *Actinidia* NR was assigned the nearest *Arabidopsis* BLASTx match (E < 1.0e-10) if there was no conflicting evidence from domains or families detected by InterProScan. The NR and its EST members were then assigned a Mapman classification bin and bin name based on that assigned to the *Arabidopsis* match.

Homology searching and phylogenetic analysis

Previously published protein sequences from GenBank were used to identify genes in the *Actinidia* EST database using tBLASTn searches [93], typically with a cutoff value of E < 1.0e-20. Identified genes were then manually checked to ensure accuracy. Amino acid alignments of predicted proteins were constructed using Clustal X. Phylogenetic analysis was carried out using the PHYLIP suite of programs [94]. Distances were calculated using protdist, and the Fitch method was used to construct a tree. Bootstrap analysis was conducted using 1000 bootstrap replicates using seqboot [94]. Treeview (v.1.6.6) was used to display resulting trees [95].

Microarray construction

For each predicted NR in the *Actinidia* EST database, a 45–55 mer oligonucleotide was created, using PERL scripts within BioView that selected all possible oligonucleotides within a gene based on uniqueness of sequence, lack of repetitive regions, and a constant melting temperature. For each NR the most 3' oligonucleotide that passed these criteria was selected for synthesis. As the EST libraries were sequenced from different *Actinidia* species, the resulting oligonucleotides were derived from genes from different *Actinidia* species. A total of 17,472 oligonucleotides were made, with the majority coming from *A. chinensis* and *A. deliciosa*. A summary of the oligonucleotides used to construct the *Actinidia* microarray slides by library and species is given in Additional file 4. Genomic DNA from *A. deliciosa* 'Hayward' and *A. eriantha* (genotype 11-6-15e) was labeled and each microarray hybridized as described previously [12]. Data were normalized using quantile normalization in the Bioconductor package Limma [96].

Chemical analysis of actinidia flowers and fruit

Volatile flavor and aroma compounds

Both headspace and solvent sampling was used to obtain volatiles. For headspace sampling of flowers, whole flowers (2.5–5 g) that were 50–75% open were placed in a 250 ml Quickfit™ Erlenmeyer flask to which was fitted a volatile trapping cartridge packed with 100 mg of Chromo-

sorb 105™ absorbent and sampled and analyzed as described earlier [29]. For fruit, ten fruit (eating-ripe) were cut longitudinally in half and one half of each fruit was used for headspace sampling and analysis [29]. The other halves were frozen in liquid N_2 for solvent extraction.

For solvent extraction, either ~1 g of flower petals were rinsed in 2 ml of purified 50:50 pentane/Et_2O or 20 g of fruit was pulped, and gently shaken in a sealed test tube with 10 mL of pentane:Et_2O (8:2) for 1–2 min, four times over a 2 h period at room temperature. Samples were processed and analyzed by GC-MS as described earlier [29].

Quantification of compounds was carried out using an average FID response based on methyl butanoate, ethyl butanoate, hexanol and methyl benzoate. Component identification was by comparison with spectra in the Mass Spectral Database (1998 NIST and an in-house database), retention indices (in-house database) and in some cases direct GC-MS comparison with authentic standards.

Analysis of fruit acids and sugars
Tissue was ground in liquid N_2 and acids and sugars were extracted from a known mass of tissue (1.4–2.5 g) into 10 mL of 80% EtOH (with adonitol and tartaric acid added as internal standards) at 60°C for 1 h. The extracts were processed, sugars and organic acids derivatized and then quantified by GC [97]. Identification of the acids and sugars was confirmed by GC-MS comparison with mixtures of authentic compounds.

Sequence data
Sequence data for the ESTs described in this paper can be found at GenBank under accession numbers FG396013 – FG528589.

Authors' contributions
EFW, EHAR, EJFS, GSR, JHB, KCS, MDT, MW, RMD, and RS all contributed to bioinformatics analysis. ACA, CAD, KBM, LLB, MM, MR, RDN, RGA, RPH, and SMB all undertook bioinformatics analysis and manuscript drafting. RNC undertook bioinformatics programming, HLB, and KK, did chemical analysis. AJM did chemical analysis and interpretation and manuscript writing. APG and EAM undertook design and manuscript drafting. BJJ, BN, DW, EG, KRL, MYW, RE, RW, YKY, and YYW did the EST sequencing and clone selection. ACR, ARF, LGF and MAN provided basic genetic material and analysis and manuscript drafting. WAL did manuscript drafting and bioinformatics analysis. BU contributed significantly by doing Mapman analysis and manuscript drafting. MWD and RJS did microarray analysis and interpretation. DC did SNP analysis and manuscript writing.

Additional material

Additional file 1
Additional Table 1. Expressed sequence tags (ESTs) sequenced in Angiosperm orders and families.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-351-S1.doc]

Additional file 2
Additional Table 2a. Summary of library names, descriptions and statistics for sequenced Actinidia *ESTs. Additional Table 2b.* Actinidia *libraries that were subtracted before sequencing to reduce redundant sequences.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-351-S2.doc]

Additional file 3
Additional Table 3 Codon usage calculated using 707 full-length Actinidia *cDNA sequences.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-351-S3.doc]

Additional file 4
Additional Table 4. Source of oligonucleotides used to construct the Actinidia *microarray slides by library and species*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-351-S4.doc]

Additional file 5
Additional Table 5. Highly expressed ESTs in the Actinidia *EST database.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-351-S5.doc]

Additional file 6
Additional Table 6. Most frequent InterPro families found in Actinidia *NRs.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-351-S6.doc]

Additional file 7
Additional Table 7. Volatiles, acids, and sugars in Actinidia *flowers and fruit*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-351-S7.doc]

References

1. Ferguson AR: **The need for characterisation and evaluation of germplasm: kiwifruit as an example.** *Euphytica* 2007, **154(3):**371-382.
2. Ferguson AR, Huang HW: **Genetic resources of kiwifruit: domestication and breeding.** *Hort Rev* 2007, **33:**1-121.
3. Nishiyama I: **Fruits of the *Actinidia* Genus.** In *Advances in Food and Nutrition Research Vol 52 Volume 52*. Edited by: Taylor SL. Academic Press; 2007:293-324.
4. Atkinson RG, MacRae EA: **Kiwifruit.** In *Transgenic Crops V Biotechnology in Agriculture and Forestry Volume 60*. Edited by: Pua EC, Davey MR. Berlin, Heidelberg, Germany , Springer-Verlag; 2007:329-346.
5. Sheng Y, Akesson C, Holmgren K, Bryngelsson C, Giamapa V, Pero RW: **An active ingredient of Cat's Claw water extracts: Identification and efficacy of quinic acid.** *J Ethnopharmacol* 2005, **96(3):**577-584.
6. Fei Z, Tang X, Alba RM, White JA, Ronning CM, Martin GB, Tanksley SD, Giovannoni JJ: **Comprehensive EST analysis of tomato and comparative genomics of fruit ripening.** *Plant J* 2004, **40(1):**47-59.
7. F-IPCfG: **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla.** In *Nature Volume 449*. Issue 7161 Nature Publishing Group; 2007:463-467.
8. Peng FY, Reid KE, Liao N, Schlosser J, Lijavetzky D, Holt R, Martinez Zapater JM, Jones S, Marra M, Bohlmann J, Lund ST: **Generation of ESTs in *Vitis vinifera* wine grape (Cabernet Sauvignon) and table grape (Muscat Hamburg) and discovery of new candidate genes with potential roles in berry development.** *Gene* 2007, **402(1-2):**40-50.
9. Moser C, Segala C, Fontana P, Salakhudtinov I, Gatto P, Pindo M, Zyprian E, Toepfer R, Grando MS, Velasco R: **Comparative analysis of expressed sequence tags from different organs of *Vitis vinifera L*.** *Funct Integr Genomics* 2005, **5(4):**208-217.
10. Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EHA, Allan AC, Beuning LL, Bowen JH, Gera E, Jamieson KR, Janssen BJ, Laing WA, McArtney S, Nain B, Ross GS, Snowden KC, Souleyre EJF, Walton EF, Yauk YK: **Analyses of expressed sequence tags from apple.** *Plant Physiol* 2006, **141(1):**147-166.
11. Moyle R, Fairbairn DJ, Ripi J, Crowe M, Botella JR: **Developing pineapple fruit has a small transcriptome dominated by metallothionein.** *J Exp Bot* 2005, **56(409):**101-112.
12. Schaffer RJ, Friel EN, Souleyre EJF, Bolitho K, Thodey K, Ledger S, Bowen JH, Ma JH, Nain B, Cohen D, Gleave AP, Crowhurst RN, Janssen BJ, Yao JL, Newcomb RD: **A genomics approach reveals that aroma production in apple is controlled by ethylene predominantly at the final step in each biosynthetic pathway.** *Plant Physiol* 2007, **144(4):**1899-1912.
13. Decroocq V, Fave MG, Hagen L, Bordenave L, Decroocq S: **Development and transferability of apricot and grape EST microsatellite markers across taxa.** *Theor Appl Genet* 2003, **106(5):**912-922.
14. Fraser LG, Harvey CF, Crowhurst RN, Silva HN: **EST-derived microsatellites from *Actinidia* species and their potential for mapping.** *Theor Appl Genet* 2004, **108(6):**1010-1016.
15. Fraser LG, McNeilage MA, Tsang GK, Harvey CF, De Silva H: **Cross-species amplification of microsatellite loci within the dioecious, polyploid genus *Actinidia* (Actinidiaceae).** *Theoret Appl Genet* 2005, **112(1):**149-157.
16. Kantety RV, La Rota M, Matthews DE, Sorrells ME: **Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat.** *Plant Mol Biol* 2002, **48(5-6):**501-510.
17. Lindqvist C, Scheen AC, Yoo MJ, Grey P, Oppenheimer DG, Leebens-Mack JH, Soltis DE, Soltis PS, Albert VA: **An expressed sequence tag (EST) library from developing fruits of an Hawaiian endemic mint (*Stenogyne rugosa*, Lamiaceae): characterization and microsatellite markers.** *BMC Plant Biol* 2006, **6(1):**16.
18. Rossetto M, McNally J, Henry RJ: **Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae.** *Theor Appl Genet* 2002, **104(1):**61-66.
19. Yu JK, La Rota M, Kantety RV, Sorrells ME: **EST derived SSR markers for comparative mapping in wheat and rice.** *Mol Genet Genomics* 2004, **271(6):**742-751.
20. Zhang L, Yuan D, Yu S, Li Z, Cao Y, Miao Z, Qian H, Tang K: **Preference of simple sequence repeats in coding and non-coding**

regions of *Arabidopsis thaliana*. *Bioinformatics* 2004, **20:**1081-1086.
21. Praekelt UM, McKee RA, Smith H: **Molecular analysis of actinidin, the cysteine proteinase of *Actinidia chinensis*.** *Plant Mol Biol* 1988, **10(3):**193-202.
22. Nieuwenhuizen NJ, Beuning LL, Sutherland PW, Sharma NN, Cooney JM, Bieleski LRF, Schröder R, MacRae EA, Atkinson RG: **Identification and characterisation of acidic and novel basic forms of actinidin, the highly abundant cysteine protease from kiwifruit.** *Funct Plant Biol* 2007, **34(10):**946-961.
23. Malone LA, Todd JH, Burgess EPJ, Philip BA, Christeller JT: **Effects of kiwifruit (*Actinidia deliciosa*) cysteine protease on growth and survival of *Spodoptera litura* larvae (*Lepidoptera: Noctuidae*) fed with control or transgenic avidin-expressing tobacco .** *NZ J Crop Hort Sci* 2005, **33:**99–105.
24. Ledger SE, Gardner RC: **Cloning and characterization of five cDNAs for genes differentially expressed during fruit development of kiwifruit (*Actinidia deliciosa var. deliciosa*).** *Plant Mol Biol* 1994, **25(5):**877-886.
25. Tamburrini M, Cerasuolo I, Carratore V, Stanziola AA, Zofra S, Romano L, Camardella L, Ciardiello MA: **Kiwellin, a novel protein from kiwi fruit. Purification, biochemical characterization and identification as an allergen.** *Protein J* 2005, **24(7-8):**423-429.
26. Davies C, Robinson SP: **Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterization of cDNAs encoding putative cell wall and stress response proteins.** *Plant Physiol* 2000, **122:**803-812.
27. Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M: **MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.** *Plant J* 2004, **37(6):**914-939.
28. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Buillard V, Cerutti L, Copley R, Courcelle E, Das U, Daugherty L, Dibley M, Finn R, Fleischmann W, Gough J, Haft D, Hulo N, Hunter S, Kahn D, Kanapin A, Kejariwal A, Labarga A, Langendijk-Genevaux PS, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Nikolskaya AN, Orchard S, Orengo C, Petryszak R, Selengut JD, Sigrist CJA, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C: **New developments in the InterPro database.** *Nucleic Acids Res* 2007, **35():**D224-D228.
29. Matich AJ, Young H, Allen JM, Wang MY, Fielder S, McNeilage MA, MacRae EA: ***Actinidia arguta*: volatile compounds in fruit and flowers.** *Phytochemistry* 2003, **63(3):**285-301.
30. Hyeon SB, Isoe S, Sakan T: **The structure of neomatatabiol, the potent attractant for Chrysopa from *Actinidia polygama* MIQ.** *Tet Lett* 1968, **51:**5325-5326.
31. Yoshihara K, Sakai T, Sakan T: **Dehydroiridodial, the pungent principle of *Actinidia polygama* Miq.** *Chem Lett* 1978, **7:**433-434.
32. Burdock GA: **Fenaroli's Handbook of Flavor Ingredients.** Boca Raton , CRC Press; 2005.
33. Friel EN, Wang M, Taylor AJ, MacRae EA: ***In vitro* and *in vivo* release of aroma compounds from yellow-fleshed kiwifruit.** *J Agric Food Chem* 2007, **55(16):**6664-6673.
34. Paterson VJ, MacRae EA, Young H: **Relationships between sensory properties and chemical composition of kiwifruit (*Actinidia deliciosa*).** *J Sci Food Agric* 1991, **57:**235-251.
35. Shaw GJ, Allen JM, Yates MK, Franich FA: **Volatile flavour constituents of feijoa (*Feijoa sellowiana*) - analysis of fruit flesh.** *J Sci Food Agric* 1990, **50(3):**357-361.
36. Flath RA, Black DR, Guadagni DG, McFadden WH, Schultz TH: **Identification and organoleptic evaluation of compounds in Delicious apple essence .** *J Agric Food Chem* 1967, **15(1):**29-35.
37. Vollhardt KPC, Schore NE: **Organic Chemistry.** Second edition. New York , W. H. Freeman and Company; 1994.
38. Akashi T, Aoki T, Ayabe S: **Molecular and biochemical characterization of 2-hydroxyisoflavanone dehydratase. Involvement of carboxylesterase-like proteins in leguminous isoflavone biosynthesis.** *Plant Physiol* 2005, **137(3):**882-891.
39. Pontier D, Godiard L, Marco Y, Roby D: **hsr203J, a tobacco gene whose activation is rapid, highly localized and specific for incompatible plant/pathogen interactions.** *Plant J* 1994, **5(4):**507-521.

40. Ueguchi-Tanaka M, Nakajima M, Katoh E, Ohmiya H, Asano K, Saji S, Hongyu X, Ashikari M, Kitano H, Yamaguchi I, Matsuoka M: **Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellin.** *Plant Cell* 2007, **19(7):**2140-2155.

41. Marshall SD, Putterill JJ, Plummer KM, Newcomb RD: **The carboxylesterase gene family from *Arabidopsis thaliana*.** *J Mol Evol* 2003, **57(5):**487-500.

42. Ileperuma NR, Marshall SD, Squire CJ, Baker HM, Oakeshott JG, Russell RJ, Plummer KM, Newcomb RD, Baker EN: **High-resolution crystal structure of plant carboxylesterase AeCXE1, from *Actinidia eriantha*, and its complex with a high-affinity inhibitor paraoxon.** *Biochem* 2007, **46(7):**1851-1859.

43. Lange BM, Rujan T, Martin W, Croteau R: **Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes.** *Proc Natl Acad Sci USA* 2000, **97(24):**13172-13177.

44. Green S, Friel EN, Beuning L, Matich A, MacRae E: **A multifunctional germacrene D synthase. Patent # WO04058814A1.** 2004.

45. McGhie TK, Ainge GD: **Color in fruit of the genus *Actinidia*: carotenoid and chlorophyll compositions.** *J Agric Food Chem* 2002, **50(1):**117-121.

46. Hörtensteiner S: **Chlorophyll degradation during senescence.** *Ann Rev Plant Biol* 2006, **57:**55-77.

47. Cunningham FX, Gantt E: **Genes and enzymes of carotenoid biosynthesis in plants.** *Ann Rev Plant Physiol and Plant Mol Biol* 1998, **49(1):**557-583.

48. DellaPenna D, Pogson BJ: **Vitamin synthesis in plants: tocopherols and carotenoids.** *Ann Rev Plant Biol* 2006, **57(1):**711-738.

49. Grotewold E: **The genetics and biochemistry of floral pigments.** *Annu Rev Plant Biol* 2006, **57:**761-780.

50. Cunningham FX, Pogson B, Sun Z, McDonald KA, DellaPenna D: **Functional analysis of the beta and epsilon cyclase enzymes of Arabidopsis reveals a mechanism for control of cyclic carotenoid formation.** *Plant Cell* 1996, **8:**1613-1626.

51. Kim J, DellaPenna D: **Defining the primary route for lutein synthesis in plants: The role of Arabidopsis carotenoid beta-ring hydroxylase CYP97A3.** *Proc Natl Acad Sci USA* 2006, **103(9):**3474-3479.

52. Ikegami A, Eguchi S, Kitajima A, Inoue K, Yonemori K: **Identification of genes involved in proanthocyanidin biosynthesis of persimmon (*Diospyros kaki*) fruit.** *Plant Sci* 2007, **172(5):**1037-1047.

53. Montefiori M, McGhie TK, Costa G, Ferguson AR: **Pigments in the Fruit of Red-Fleshed Kiwifruit (*Actinidia chinensis* and *Actinidia deliciosa*).** *J Agric Food Chem* 2005, **53(24):**9526-9530.

54. Ferguson AR, MacRae EA: **Vitamin C in *Actinidia*.** In *Acta Hortic Volume 297*. Wageningen , International Society for Horticultural Science; 1992:481-487.

55. Ishikawa T, Dowdle J, Smirnoff N: **Progress in manipulating ascorbic acid biosynthesis and accumulation in plants.** *Physiol Plant* 2006, **126(3):**343-355.

56. Boldingh H, Smith GS, Klages K: **Seasonal concentrations of non-structural carbohydrates of five *Actinidia* species in fruit, leaf and fine root tissue.** *Ann Bot* 2000, **85(4):**469-476.

57. Klages K, Donnison H, Boldingh H, MacRae E: **myo-Inositol is the major sugar in *Actinidia arguta* during early fruit development.** *Aust J Plant Physiol* 1998, **25(1):**61 -668.

58. Laing WA, Wright MA, Cooney J, Bulley SM: **The missing step of the L-galactose pathway of ascorbate biosynthesis in plants, an L-galactose guanyltransferase, increases leaf ascorbate content.** *Proc Natl Acad Sci USA* 2007, **104(22):**9534-9539.

59. Walton EF, De Jong TM: **Growth and compositional changes in kiwifruit berries from three Californian locations.** *Ann Bot* 1990, **66:**285-298.

60. Cheng C, Seal A, Boldingh H, Marsh K, MacRae EA, Murphy S, Ferguson AR: **Inheritance of fruit characters and fruit size in a diploid *Actinidia Chinensis* (kiwifruit) population.** *Euphytica* 2004, **138:**185-195.

61. Jensen HD, Krogfelt KA, Cornett C, Hansen SH, Christensen SB: **Hydrophilic carboxylic acids and iridoid glycosides in the Juice of American and European Cranberries (*Vaccinium macrocarpon* and *V. oxycoccos*), Lingonberries (*V. vitis-idaea*), and Blueberries (*V. myrtillus*).** *J Agric Food Chem* 2002, **50(23):**6871-6874.

62. Dirlewanger E, Moing A, Rothan C, Svanella L, Pronier V, Guye A, Plomion C, Monet R: **Mapping QTLs controlling fruit quality in peach (*Prunus persica* (L.) Batsch).** *Theoret Appl Genet* 1999, **98(1):**18-31.

63. Ossipov V, Bonner C, Ossipova S, Jensen R: **Broad-specificity quinate (shikimate) dehydrogenase from *Pinus taeda* needles.** *Plant Physiol Biochem* 2000, **38(12):**923-928.

64. Sampson HA: **Food allergy - accurately identifying clinical reactivity.** *Allergy* 2005, **60(Suppl 79):**19-24.

65. Lucas JS, Lewis SA, Hourihane JO: **Kiwi fruit allergy: a review.** *Pediatr Allergy Immunol* 2003, **14:**420–428..

66. Lucas JS, Grimshaw KE, Collins K, Warner JO, Hourihane JO: **Kiwi fruit is a significant allergen and is associated with differing patterns of reactivity in children and adults.** *Clin Exp Allergy* 2004, **34(7):**1851-1855.

67. Chen L, Lucas JS, Hourihane JO, Lindemann J, Taylor SL, Goodman RE: **Evaluation of IgE binding to proteins of hardy (*Actinidia arguta*), gold (*Actinidia chinensis*) and green (*Actinidia deliciosa*) kiwifruits and processed hardy kiwifruit concentrate, using sera of individuals with food allergies to green kiwifruit.** *Food Chem Toxicol* 2006, **44(7):**1100-1107.

68. Mills CE, Shewry PR: **Plant food allergens.** Oxford , Blackwell Science ; 2004.

69. Pastorello EA, Conti A, Pravettoni V, Farioli L, Rivolta F, Ansaloni R, Ispano M, Incorvaia C, Giuffrida MG, Ortolani C: **Identification of actinidin as the major allergen of kiwi fruit.** *J Allergy Clin Immunol* 1998, **101(4 Pt 1):**531-537.

70. Moller M, Kayma M, Steinhart H, Paschke A: **Isolation and characterization of a major allergen in kiwi fruit.** *Z Lebensm Unters Forsch* 1997, **205:**364–369.

71. Breiteneder H: **Thaumatin-like proteins -- a new family of pollen and fruit allergens.** *Allergy* 2004, **59(5):**479-481.

72. Robert LS, Nozzolillo C, Altosaar I: **Homology between legumin-like polypeptides from cereals and pea.** *Biochem J* 1985, **226(3):**847-852.

73. Singh NK, Shepherd KW, Langridge P, Gruen LC: **Purification and biochemical characterization of triticin, a legumin-like protein in wheat endosperm.** *J Cereal Sci* 1991, **13(3):**207-219.

74. Duranti M, Horstmann C, Gilroy J, Croy RR: **The molecular basis for N-glycosylation in the 11S globulin (legumin) of lupin seed.** *J Protein Chem* 1995, **14(2):**107-110.

75. Gruis D, Selinger DA, Curran JM, Jung R: **Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases.** *Plant Cell* 2002, **14(11):**2863-2882.

76. Muntz K: **Globulins from legume seeds: Structure and function during storage and reactivation.** In *Plant proteins from European crops : food and non-food applications* Edited by: Gueguen J, Popineau Y. Berlin ; New York , Springer, c1998. xxvi, 339 p. : ill.; 1998:3-12.

77. Shutov AD, Vaintraub IA: **Degradation of storage proteins in germinating seeds.** *Phytochem* 1987, **26(6):**1557-1566.

78. White A, de Silva HN, Requejo-Tapia C, Harker FR: **Evaluation of softening characteristics of fruit from 14 species of Actinidia.** *Postharv Biol and Tech* 2004, **35:**143–151..

79. Jackson PJ, Harker FR: **Changes in the firmness of the outer pericarp, inner pericarp, and core of Actinidia species during ripening.** *NZ J Crop Hort Sci* 1997, **25:**185–189.

80. Schröder R, Atkinson RG: **Kiwifruit cell walls: towards an understanding of softening?** *NZ J For Sci* 2006, **36:**112-129.

81. Brummell DA, Harpster MH: **Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants.** *Plant Mol Biol* 2001, **47(1-2):**311-340..

82. Smith DL, Starrett DA, Gross KC: **A gene coding for tomato fruit beta -Galactosidase II is expressed during fruit ripening . Cloning, characterization, and expression pattern.** *Plant Physiol* 1998, **117(2):**417-423.

83. Ross GS, Redgwell RJ, MacRae EA: **Kiwifruit β-galactosidase: Isolation and activity against specific fruit cell-wall polysaccharides.** *Planta* 1993, **189:**499–506.

84. Schröder R, Atkinson RG, Langenkämper G, Redgwell RJ: **Biochemical and molecular characterisation of xyloglucan endotransglycosylase from ripe kiwifruit.** *Planta* 1998, **204:**242-251.

85. Gill GP, Harvey CF, Gardner RC, Fraser LG: **Development of sex-linked PCR markers for gender identification in Actinidia.** *Theoret Appl Genet* 1998, **97(3):**439-445.

86. Nakamura Y, Gojobori T, Ikemura T: **Codon usage tabulated from international DNA sequence databases: status for the year 2000.** *Nucl Acids Res* 2000, **28(1):**292.
87. Chang S, Puryear J, Cairney J: **A simple and efficient method for isolating RNA from pine trees.** *Plant Mol Biol Rep* 1993, **11(2):**113-116.
88. Lopez-Gomez R, Gomez-Lim MA: **A method for extracting intact RNA from fruits rich in polysaccharides using ripe mango mesocarp.** *HortSci* 1992, **27(5):**440-442.
89. Zdobnov EM, Apweiler R: **InterProScan--an integration platform for the signature-recognition methods in InterPro.** *Bioinformat* 2001, **17(9):**847-848.
90. Garcia-Hernandez M, Berardini TZ, Chen G, Crist D, Doyle A, Huala E, Knee E, Lambrecht M, Miller N, Mueller LA, Mundodi S, Reiser L, Rhee SY, Scholl R, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P: **TAIR: a resource for integrated Arabidopsis data.** *Funct Integr Genomics* 2002, **2(6):**239-253.
91. Huang X, Madan A: **CAP3: A DNA sequence assembly program.** *Genome Res* 1999, **9(9):**868-877.
92. Rice P, Longden I, Bleasby A: **EMBOSS: The European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16:**276–277.
93. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3):**403-410.
94. Felsenstein J: **PHYLIP (Phylogeny Inference Package), Version 3.6a2.** 3.6a2 edition. 1993 [http://evolution.genetics.washington.edu/phylip.html]. Department of Genetics, University of Washington, Seattle.
95. Page RDM: **Treeview: An application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12(4):**357-358.
96. Smyth GK, Speed TP: **Normalization of cDNA microarray data.** *Methods* 2003, **31:**265-273.
97. Richardson AC, Marsh KB, Boldingh HL, Pickering AH, Bulley SM, Frearson NJ, Ferguson AR, Thornber SE, Bolitho KM, Macrae EA: **High growing temperatures reduce fruit carbohydrate and vitamin C in kiwifruit.** *Plant, Cell & Environment* 2004, **27(4):**423-435.
98. Herrmann KM, Weaver LM: **The shikimate pathway.** *Ann Rev Plant Physiol Plant Mol Biol* 1999, **50:**473-503.
99. Singh SA, Christendat D: **Structure of Arabidopsis dehydroquinate dehydratase-shikimate dehydrogenase and implications for metabolic channeling in the shikimate pathway.** *Biochemistry* 2006, **45(25):**7787-7796.
100. Urbanczyk-Wochniak E, Usadel B, Thimm O, Nunes-Nesi A, Carrari F, Davy M, Bläsing O, Kowalczyk M, Weicht D, Polinceusz A, Meyer S, Stitt M, Fernie A: **Conversion of MapMan to allow the analysis of transcript data from Solanaceous species: effects of genetic and environmental alterations in energy metabolism in the leaf.** *Plant Mol Biol* 2006, **60(5):**773-792.