A Diploid Platform for Strawberry Genomics

Davis TM, Shields ME, DiMeglio LM, Brese RL, and Zhang Q
Department of Plant Biology and Genetics Graduate Program
University of New Hampshire, Durham NH

Concept

The ability of strawberry researchers and breeders to address the present and future needs of the United States strawberry industry will be enhanced by knowledge of strawberry genome structure/function and access to appropriate molecular tools. Although the small (~200 Mb) size of the basic (x=7) strawberry genome is highly favorable, the octoploid (2n=8x=56) genome structure and allelic complexity of the cultivated strawberry, *Fragaria ×ananassa*, is a complicating factor for both structural and functional genomics. A solution is to use appropriate diploid (2n=2x=14) *Fragaria* species as a research platform for genomic resource development.

The octoploid strawberry's proposed genome composition of AAA'A'BBB'B' implies subgenome contributions from as many as four different diploids. An initial phylogenetic analysis provides evidence that *Fragaria vesca* and *F. iinumae* were likely genome donors. The ready availability of suitable germplasm and prior knowledge and resource development justifies a focus on *Fragaria vesca* as an initial subject for genomic resource development, but equal attention to *F. iinumae* is needed.

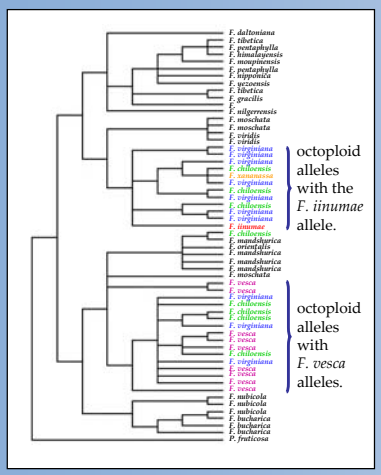
Project Goal

To establish structural and functional genomic resources that could be of immediate utility for basic and applied strawberry genetic research and germplasm development.

Project Objectives

Are centered on 1) developing structural genomics resources, including a large insert genomic library; and 2) constructing a cDNA library from *F. vesca* floral tissue and expanding the strawberry EST database.

Preliminary Phylogenetic Analysis - ADH Intron Sequence -



Germplasm Collection

F. iinumae

The geographic distribution of *F. iinumae* is limited to Japan and narrowly adjacent mainland. As a step toward future investigations, separate funding was obtained from the USDA Germplasm System (Kim Hummer, P.I.) to expand the available collection of *F. iinumae* (18 new accessions in NCGR) and other fruit germplasm from the island of Hokkaido, Japan.

Significance: Provides foundation for needed development of genomic resources derived from a second *Fragaria* genome.

Fragaria vesca: A Diploid Model Species for Strawberry

F. vesca is the most widely distributed and diverse of the *Fragaria* species, occurring in a circumpolar distribution and having four subspecies: *vesca*, *americana*, *bracteata*, and *californica*. Derivatives of ssp. *vesca* have been cultivated in Europe for hundreds of years, with emphasis on the 'Alpine' or 'Semperflorens' (perpetual flowering) form.

F. vesca is self-compatible, has a ~4 month seed-to-seed generation time, and can be easily propagated vegetatively and in vitro. The 'Alpine' forms are perpetual flowering, as opposed to the short day photoperiod sensitivity of most wild type forms. Several other recessive mutant traits occur in some 'Alpine' varieties, including yellow fruit, runnerless, and others. These varieties have been inbred over many generations through varietal seed propagation.

Our investigations have two employed 'Alpine' varieties, 'Baron Solemacher' and 'Yellow Wonder' and two representatives of ssp. *americana*, WC6 and 'Pawtuckaway', collected from the wild in New Hampshire.

F. vesca MAPPING POPULATIONS & MAPS*

---------- PARENTS ----------

ssp. *vesca* 'Alpine'	ssp. *americana*
'Baron Solemacher' (BS)	WC6
'Yellow Wonder' (YW)	'Pawtuckaway' (P)

---------- CROSSES AND F2 MAPPING POPULATIONS ----------

BS x WC6 (N = 80)

Basis for first Fragaria linkage map (Davis and Yu, 1997). Plants no longer exist, but markers still being added using archived DNA samples.

YW x P (N = 113)

Basis for second generation map (in progress). Mapping population maintained at UNH; additional F2 plants at IRTA, Cabrils, Spain. Segregating for perpetual flowering, runnerless, yellow fruit, other traits

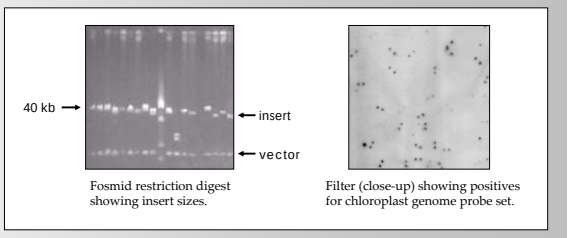
**Note: Dan Sargent's diploid reference map is based upon an interspecific cross: F. vesca x F. nubicola.*

Large Insert Genomics Library

Approach: A fosmid library was constructed from *F. vesca* ssp *americana* 'Pawtuckaway', using the Epicentre CopyControl pCC1FOS vector system.

Rationale: Physically sheared DNA and blunt-end cloning provide optimal genomic representation. Pawtuckaway chosen as source of "wild type" alleles.

Outcome: ~33,000 clones picked, and spotted onto filter sets. Insert size range 30-40 kb. 6x nuclear genome coverage. Organelle genome content ~ 11%.



Fosmid restriction digest showing insert sizes.

Filter (close-up) showing positives for chloroplast genome probe set.

Initial Sequence Sample

Approach: This fosmid was selected using an alcohol dehydrogenase (ADH) gene probe.

Annotation utilized FGENESH, Blastn, Blastx, to EST, NR and other genomic databases, SSRIT.

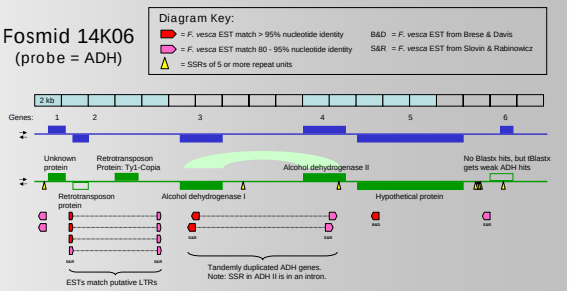
Credit: Subcloning and sequencing performed by Ana Pontaroli in Jeff Bennetzen's lab.

FGENESH
Arabidopsis

Blastx-nr
Viridiplantae

Blastn-EST
Fragaria

Fosmid 14K06
(probe = ADH)



Key Findings:

The targeted ADH gene, previously thought to be single copy in *F. vesca*, was found to be tandemly duplicated. The previously known copy we designate ADH I, while the newly discover gene is designated ADH II. A third, cryptic ADH gene occurs at the right end of the fosmid. The two *Fragaria* ESTs that match ADH correspond best to ADH I.

Four ESTs (Slovin and Rabinovicz) may correspond to the LTR regions of a retrotransposon. Paired hits for these EST, each defining a 7 kb interval, have been found on four of 50 *F. vesca* fosmids.

This and other results served as preliminary data in a successful USDA-NRI grant proposal, the outcome of which is described on the accompanying poster.

EST Resource Development

Rationale: At the time this project was initiated, there were fewer than 1000 *Fragaria* ESTs in GenBank.

Flower buds are anatomically complex, providing a highly diverse population of transcripts. Flowering-related traits are of broad interest.

Use of an inbred variety 'Yellow Wonder' minimized allelic diversity, simplifying analysis.

Approach: A flower bud cDNA library was constructed from *F. vesca* ssp. *vesca* 'Yellow Wonder', using the CLONTECH SMART cDNA Library Construction Kit.

Unopened flower buds, with sepals removed to minimize contaminants, were used as RNA source.

Assistance with bioinformatic analysis was provided by Dorrie Main and Meg Staton (Genome Database for Rosaceae = GDR).
~800 cDNA clones sequenced courtesy of V. Shulaev, VBI.

Outcome: About 3,300 clones sequenced. 2717 high quality EST sequences were deposited into GenBank, under accession #s DV438013-DV440734.

There were ~1900 unigenes (unique singlets plus contigs).

148 unigenes had SSRs of five or more repeat units.

Significance & Deliverables

- YW x P mapping population – basis for new *F. vesca* map.
- *F. vesca* fosmid library - an invaluable genomic resource. Basis for subsequent USDA-NRI Project (see accompanying poster).
  - Fosmid clones provided to other labs: Abbott, JGI
  - Fosmid filters provided to other labs: Abbott, Slovin.
- Flower bud cDNA library – only *Fragaria* floral library.
  - cDNA plates shared with Timo Hytönen lab, U. Helsinki.
- Over 25% of flower bud ESTs still unique in *Fragaria* EST database, suggesting value and need for additional floral cDNA cloning and sequencing.
- These and other *Fragaria* ESTs are an invaluable resources for use in basic research, marker development, reverse genetics, and annotation of genomic sequence.

Acknowledgements:

This research is supported by USDA-CSREES NRI Plant Genome Grant 2003-35300-3142, and New Hampshire Agricultural Experiment Station Project NH00433.

Sequence Samples & Gene Pair Haplotypes in Strawberry

Davis TM, Shields ME, and Zhang Q
Department of Plant Biology & Genetics Graduate Program
University of New Hampshire, Durham NH

Bennetzen JL and Pontaroli AC
Department of Genetics
University of Georgia, Athens GA

Tombolato DCM and Folta KM
Horticultural Sciences Department
University of Florida, Gainesville FL

San Miguel P
Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette IN

Background

The cultivated strawberry, *Fragaria ×ananassa* is an octoploid (2N = 8X = 56), hybrid species, arising less than 300 years ago via hybridization between octoploids *F. chiloensis* and *F. virginiana*. The hypothesized genome composition of the octoploid strawberries is AAA'A'BBB'B', with subgenome contributions from up to four different diploids. An initial phylogenetic analysis draws attention to four diploids (2N = 2X = 14) as possible genome donors: primary candidates *F. vesca* and *F. iinumae* and secondary candidates *F. mandshurica* and *F. bucharica*.

Considerable attention has been devoted to development of *F. vesca* as a diploid model species for *Fragaria*. Its favorable features include small (~200 Mb) genome, self-compatibility, short generation time, small plant size, ease of seed and vegetative propagation, availability of useful mutants and inbred lines, and ease of genetic transformation.

Project Overview

We have constructed an *F. vesca* genomic library in a fosmid vector using physically sheared DNA, and have sequenced 20 gene-targeted and 30 randomly selected clones, generating a total of 1.75 Mb genomic sequence. The 50 sequences are deposited under GenBank accession numbers EU024823-EU024872. These sequences have been exploited in the development of two new marker concepts – gene pairs, and gene pair haplotypes.

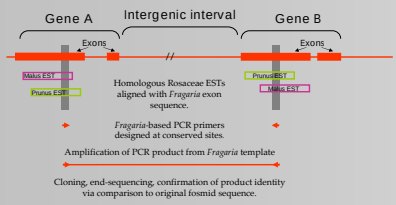
Gene Pair Markers

In the present context, the term "gene pair" refers to two immediately adjacent genes. The placement of a "forward" PCR primer site in a proximal exon in gene A, and the "reverse" primer site in a proximal exon of immediately adjacent gene B allows amplification of the A-B gene pair interval, which includes the intergenic region as well as any exon (and intron) sequences flanked by the PCR primers. Intergenic regions are rich in polymorphisms, including SNPs, indels and SSRs.

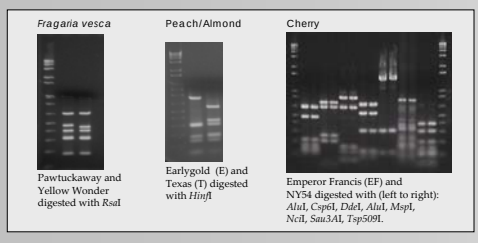
The gene pair marker approach exploits this abundant polymorphism to achieve convenient differentiation between mapping parents, and genotyping of segregating populations, using the CAPS (cleaved amplified polymorphic site) technique - i.e., by restriction enzyme digestion of PCR products followed by visualization on an ethidium bromide stained electrophoretic gel, or on a sequencing gel in conjunction with fluorescent labeling.

Marker Development

Approach: *Fragaria* genomic sequence (from fosmid clone) defines gene pair site. Primers are designed at exon sites conserved in Rosaceae, based on EST comparisons. Products are cloned and sequenced for verification of identity. Species-specific primers are developed as needed.



Example: CDPK-BHLH Gene Pair, application to *F. vesca* and transfer to other rosaceous species. Genotyping employs restriction digestion of PCR products (CAPS technique).



Outcome: ~2 Kb products from *F. vesca*, rose, and cherry were cloned and end-sequenced. In each case product identities were confirmed, validating gene pair site, and demonstrating its conservation among these species. Polymorphism detection in each species demonstrates marker transferability. This gene pair locus, which is on *F. vesca* linkage group I, was mapped onto *Prunus* LG 7, in cooperation with Werner Howad, IRTA, using the *Prunus* TxE BinMap set, illustrating the potential use of gene pair markers for comparative mapping among Rosaceae genera.

Features of *F. vesca* Fosmid Clones

Key Findings: In a sampling of 1.75 Mb genomic sequence distributed over 50 sites, of which 30 were selected at random, gene density was about 1 protein-encoding gene per 6 kb.

Fosmids are being placed on the *F. vesca* map using gene pair markers.

Fragaria ESTs were vital resources for gene validation and annotation.

Several genes were tandemly duplicated, included the targeted genes ADH, CHS, Pistillata, terpene synthase, and cRGA1.

SSRs of five or more units also occurred at a ~1 per 6 kb density.

Gypsy and Copia retrotransposon-like sequences were found on several fosmids.

Microcolinearity, involving two to five genes, was common in comparisons with *Arabidopsis*, *Populus*, *Vitis*, and others.

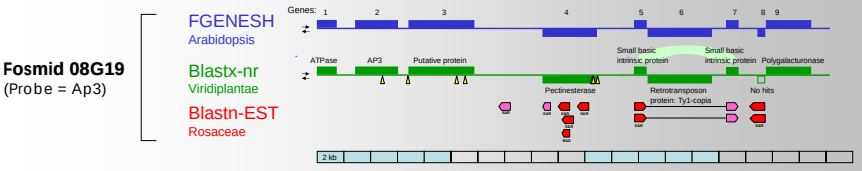
Fosmid Annotation: FGENESH, Blastn, Blastx, to EST, NR and other genomic databases. SSRIT.

Gene-targeted Fosmids

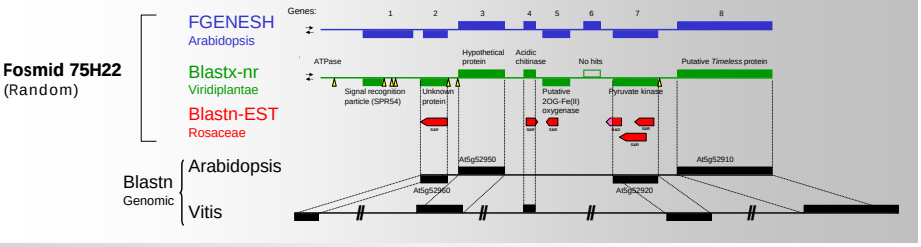
	Fosmid	GenBank	~ kb*	Probe	LG	# SSRs	Notes
Anthocyanin biosynthesis	48I08	EU024856	39	CHI	VII	11	
	53O08	EU024865	31	DFR	II	8	Retroelement
	73I22	EU024868	33	CHS	VII	4	Tandem gene duplication
Other metabolism	76C08	EU024870	45	RAN	I	16	Retroelement
	14K06	EU024832	36	ADH	II	6	Tandem gene duplication
	52I20	EU024863	42	GBSS1	VII	10	
	41O22	EU024852	33	Terp. Syn.	III	5	Tandem gene duplication
	53J04	EU024864	33	Terp. Syn.	III	6	Retroelement
Flowering related	08G19	EU024826	38	Ap3	VI	7	Retroelement Ty1-copia
	10B08	EU024827	35	Leafy	III	13	
	76K13	EU024871	33	Pistillata	II	5	Tandem gene duplication
	51F10	EU024860	30	PHYA	VI	8	
	01I13	EU024823	41	Hy5	II	7	
	52B01	EU024861	30	Constans	VI	4	
Disease resistance	19H07	EU024837	32	SOC	VII	4	
	19M24	EU024838	33	PRGA1	IV	2	
	32L07	EU024845	33	PRGA2	IV	0	
	34E24	EU024847	37	cRGA1	I	12	Tandem gene duplication
	13I24	EU024831	34	PRGA2		8	
	49M15	EU024859	42	Ve-like		19	Retroelement

Diagram Key:
= *F. vesca* EST match > 95% nucleotide identity
= *F. vesca* EST match 80 - 95% nucleotide identity
= SSRs of 5 or more repeat units
B&D = *F. vesca* EST from Brese & Davis
S&R = *F. vesca* EST from Slovin & Rabinowicz

Fosmid 08G19 (Probe = Ap3)



Fosmid 75H22 (Random)

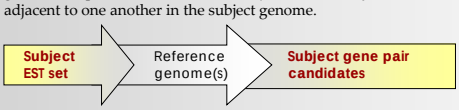


Blastn Genomic: Arabidopsis, Vitis

Gene Pair Detective

For extending gene pair markers to species lacking adequate genomic sequence databases, a computational tool named "Gene Pair Detective" was designed (in cooperation with Viplav Mishra and Tamer Kahveci, U. Florida) to compare EST sets from under-sequenced genomes against the genomes of reference organisms.

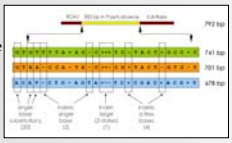
The Gene Pair Detective program implements Blastn to populate an M x N matrix, and a bipartite matching algorithm defines the best match. The program exploits the validated possibility of locally conserved gene order between subject and reference genomes to predict ESTs from the subject set that may reside adjacent to one another in the subject genome.



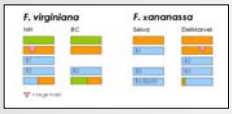
Gene Pair Haplotypes

In the octoploid strawberry, as many as eight different alleles could exist at a given locus, and 16 alleles could be segregating in a cross between maximally heterozygous individuals. Single site markers are ill-suited for resolving such allelic diversity, but multi-character haplotypes may suffice. Gene pair markers target intergenic regions rich in sequence polymorphisms of various types, as needed to define multi-character haplotypes.

Example:
At the RGA1-Subtilase gene pair site, three major haplotypes were distinguished in four octoploids (right).



Distinct subtypes, including recombinants, were also definable, such that ten distinct haplotypes could be discerned among four varieties. From an NH x DelMarvel cross (right), an individual with eight distinct haplotypes could be derived.



Our project, currently in progress, seeks to define haplotype variation at one or more gene pair loci on each strawberry chromosome, and to assess haplotype segregation in the progeny of an octoploid. Initial simplification of allelic complexity will be achieved by employing the pentaploid progeny of a diploid x pentaploid cross.

Pentaploid Population

DIPLOID	X	OCTOPLOID
F. vesca ssp. *vesca* 'Semperflorens'		*F. virginiana* hybrid LB48

PENTAPLOID POPULATIONS

F1: 'Yellow Wonder' x LB48 (n = 25)
F1: 'Hawaii 4' x LB48 (n = 35)

Rationale: Use of diploid as tester, exposes segregation from octoploid while simplifying genotype resolution.

Deliverables & Significance

Fifty fosmid sequences deposited in GenBank under accession numbers EU024823-EU024872.

The 1.75 Mb total of 50 sequence samples is currently largest genomic database in the Rosaceae, providing an informative glimpse of genome organization in a rosaceous species, and a useful comparator for anticipated peach and apple genome sequences.

Gene pair markers are a promising new marker type, were used to map fosmids in *F. vesca*, and are applicable to species with sufficiently small (i.e., in the 1-5 kb range) intergenic distances.

Fragaria-derived gene pair markers are transferable to other Rosaceae species, providing robust anchor markers for comparative mapping.

Gene Pair Detective will facilitate development of gene pair markers in under-studied species.

With gene pair haplotypes, complete resolution of locus-specific allele content is feasible in complex polyploids, allowing definitive resolution of disomic versus polysomic inheritance patterns.

The pentaploid population will enable enhanced resolution of octoploid transmission patterns, genotyped as gene pair haplotypes.

Acknowledgements:

This research is supported by USDA-CSREES NRI Plant Genome Grant 2005-35300-15467, and New Hampshire Agricultural Experiment Station Project NH00433.