A taste of pepper: Genetics, biochemistry and prediction of sweet pepper flavor

P.M. Eggink

Thesis committee

Promotor
Prof. Dr. R.G.F. Visser
Professor of Plant Breeding, Wageningen University

Co-promotors
Dr. A.G. Bovy
Senior scientist, Plant Research International, Wageningen UR

Dr. Ir. C.A. Maliepaard
Assistant professor, Laboratory of Plant Breeding, Wageningen University

Dr. Ir. J.P.W. Haanstra
Breeding manager Pepper and Eggplant, Rijk Zwaan, De Lier

Other members
Prof. Dr. E.J. Woltering, Wageningen University
Prof. R.D. Hall, Wageningen University
Dr. P.A. Luning, Wageningen University
Dr. Ir. R.C. Schuurink, University of Amsterdam

This research was conducted under the auspices of
the Graduate School of Experimental Plant Sciences

A taste of pepper: Genetics, biochemistry and prediction of sweet pepper flavor

P.M. Eggink

Thesis
submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 6 November 2013
at 1.30 p.m. in the Aula.

P.M. Eggink
A taste of pepper:
Genetics, biochemistry and prediction of sweet pepper flavor, 160 pages.

PhD thesis, Wageningen University, Wageningen, NL (2013)
With references, with summaries in Dutch and English

ISBN 978-94-6173-724-3

CONTENTS

Chapter 1

General introduction

Consumer acceptance of vegetables is highly dependent on appearance and flavor (Rocha et al. 2013). External qualities such as color, texture and shape are relatively easy to evaluate by both producers and consumers. However, evaluation of flavor attributes is more complex. In tomato flavor research measuring physical, biochemical and sensory properties, the latter were even considered the most difficult to quantify (Fulton et al. 2002). Flavor of fruits and vegetables, as perceived during consumption has been defined as the overall sensation provided by the interaction of taste, odor, mouth feel, sight and sound. According to Luning et al. (1994b) in pepper the composition of non-volatile compounds influences mainly the sensory perceived taste, while the aroma is affected by volatile compounds. Studies about the aroma and volatile fraction of pepper (*Capsicum* spp.) fruits are, however, limited, although the interest for this topic has increased in the last decade (Moreno et al. 2012).

This chapter introduces the main topics of this thesis, including properties of the crop pepper, some aspects of fruit flavor, metabolomics and pepper genetics. Finally, the objectives and outline of the thesis are described.

Capsicum

Together with other plants such as tomato, eggplant, potato or tobacco, the genus *Capsicum* belongs to the botanical family of *Solanaceae*. Plants of the genus are some of the oldest cultivated plants in the world, as for over 6000 years their fruits have been used as spice or food in the human diet (Perry et al. 2007). The genus *Capsicum* comprises approximately 30 species (Moscone et al. 2007), which include the five domesticated species *Capsicum annuum*, *Capsicum frutescens*, *Capsicum chinense*, *Capsicum baccatum*, and *Capsicum pubescens*. Among these, *Capsicum annuum* is the most popular, genetically diverse and economically important species and its cultivars are grown worldwide (Bosland and Votava 1999, Ortiz et al. 2010). Based on molecular analyses of domesticated and wild species of *Capsicum*, it is concluded that the ancestors of *Capsicum* most likely evolved in arid regions of the Andes Mountains, in what became Peru and Bolivia, and then migrated to tropical lowland regions of the Americas (Walsh and Hoot 2001). The centers of domestication are still under discussion: *C. baccatum* and *C. pubescens* are postulated to have been domesticated in Bolivia, the putative center of crop origin of *C. annuum* is in current Mexico, and *C. chinense* and *C. frutescens* are thought to have originated from the Amazon (Moscone et al. 2007). *Capsicum* species are commonly divided into three main groups, which form the *C. annuum*, *C. baccatum* and *C. pubescens* complexes (Onus and Pickersgill, 2004). Species within the complexes can be sexually intercrossed, crosses between species from different complexes are generally more

difficult, often leading to sterile offspring or no offspring at all (Walsh and Hoot 2001).

Due to their characteristic pungency, aromas and flavors, *Capsicum* fruits, commonly known as chiles, chile pepper, ajíes, paprika, cayenne, pimiento or simply pepper, are an important ingredient in millions of people's daily diets (perhaps even billions considering India; Meckelmann et al. 2013). Peppers are eaten fresh or processed, as unripe (green or white) or ripe (e.g. red, yellow and orange) fruits. In addition, the fruits are an excellent source of health-related compounds, such as ascorbic acid (vitamin C), carotenoids (provitamin A), tocopherols (vitamin E), flavonoids and capsaicinoids (reviewed by Wahyuni et al. 2013a). The concentration and patterns of these health-promoting phytonutrients are influenced by genotype and environmental factors, as well as by processing parameters; e.g. in the production of chili powders, such as sample treatment, drying conditions, and milling (Gnayfeed et al. 2001). Next to such health-promoting attributes of pepper, its fruits are a rich source of carotenoids; capsanthin and capsorubin are the dominating carotenoids that are used for the production of natural colorants such as oleoresins, which are applied in food and cosmetics industries (Minguez-Mosquera 1998).

Flavor of fruit crops

In tomato (*Solanum lycopersicum*) the non-volatile compounds sugars and organic acids are major components of the fruits and account for ~60% of dry matter. Non-volatiles do not only contribute to soluble solids (°Brix), but both the absolute concentrations of sugars and organic acids and the balanced ratio between them are important factors in consumer acceptance of flavor (Ruiz et al. 2005). Next to non-volatile compounds, already few decades ago, more than 400 volatile compounds have been identified in tomato (Pétro-Turza 1986). Human sensitivity to such volatile chemicals varies hugely and the impact of a volatile on flavor is determined by both its concentration and its odor threshold (Baldwin et al. 2000). Based on these factors, a set of approximately 30 volatiles that significantly impact tomato flavor has been identified (Buttery 1993, Buttery and Ling 1993). Some of these volatiles have a positive impact while others are negatively perceived (Baldwin et al. 1998). Our understanding of the effect of genetic variation on these compounds is far from complete. However, an increasing attention is shifting towards exploration of genetic variation for improving crop quality (reviewed by Fernie et al. 2006).

In tomato a large number of biochemical studies have been carried out leading to QTLs that influence biochemical pathways such as those leading to sugars and organic acids. Many sugar/Brix QTLs have been defined and 63 genes that are

putatively involved in carbon metabolism have been mapped (Causse et al. 2004). A specific QTL for Brix was mapped to the cell wall invertase gene *Lin5* (Fridman et al. 2000). First steps are also made to identify the genes responsible for the synthesis of flavor related volatile compounds. Utilizing, for instance, a recombinant inbred line (RIL) population, derived from a cross between a cherry tomato line with good overall aroma intensity and an inbred line with a common taste but with bigger fruits, allowed the identification of major QTLs for six aroma volatiles (Lecomte et al. 2004). The genetically diverse but well-defined *S. pennellii* introgression line population (Eshed and Zamir, 1995) has allowed the identification of both *malodorous,* a locus affecting 2-phenyl-acetaldehyde and 2-phenylethanol (Tadmor *et al.* 2002) and 25 loci that were significantly altered in one or more of 23 different volatiles (Tieman et al. 2006). Tikunov et al. (2010) revealed the importance of volatile conjugation for storage and subsequent release of phenylpropanoid volatiles, which are responsible for smoky flavor in tomato. They recently identified a glycosyltransferase gene (*Nsgt1*), which functions as an on-off switch for the release of these volatiles (Tikunov, personal communication).

Although literature addressing flavor of some fruit crops, like tomato, strawberry, peach or melon is abundant, specific research for the fruit crop pepper is limited. Also in the breeding process of pepper, so far, other factors than flavor, like production (yield) and quality, such as shelf life, firmness and disease resistances, were of main interest. Over the last years, variety development of (mainly blocky) peppers with higher yields was at the expense of flavor (a tendency which is also widely observed in tomato, e.g. Grandillo et al. 1999). However, since consumers have become more critical, attention in pepper, like in tomato, is shifting towards flavor as a more important quality parameter (Verheul 2008). A similar tendency is seen in the growing number of recent reports addressing pepper aroma compounds. Unfortunately however, these studies are mainly limited to characterization of the variation for volatile and non-volatile compounds in cultivated and/or wild species of *Capsicum* (e.g. Kollmannsberger et al. 2011, Meckelmann et al. 2013, Wahyuni et al. 2013b), while correlations with sensory evaluations by taste or odor panels are generally missing. Few exceptions exist, e.g. research of Luning et al. (1994a) in which sensory attributes of bell peppers were correlated with the composition of volatile compounds by principal components analysis on joint gas chromatography mass spectrometry (GC-MS) and sensory data obtained from an odor panel. In this study it was found that during pepper maturation the majority of volatile compounds, of which several had green-related odor notes, decreased or even disappeared. Only the levels of 2-hexenal and 2-hexenol, which have almond, fruity, spicy and/or sweet odors, were higher at the maturity stages turning and red. Rodriguez-Burruezo et al.

(2010) isolated volatile compounds of 16 *Capsicum* accessions from the *annuum-chinense-frutescens* complex and analyzed them with combined GC-sniffing port-MS identifying more than 300 individual compounds. The sniffing test revealed that the diversity of aromas found in their panel was due to qualitative and quantitative differences of, at least, 23 odor contributing volatiles. Odors were characterized by a high contribution of esters, some ionones and the well-known bell pepper pyrazine (2-isobutyl-3-methoxypyrazine), which is commonly described as characteristic (green) bell pepper aroma (Luning et al. 1994a, Van Ruth et al. 1995). The work of Janse (1996) is another example, in which the flavor of different pepper types was characterized by taste panel evaluations and some basic biochemical analyses. In this study taste attributes 'sweet, fruity and juicy' were positively correlated to 'attractive taste'. Although sugars and organic acids were recognized to play a major role in flavor of pepper, it was concluded that the flavor of the examined varieties could only be predicted to a limited extent by measuring Brix (15% explained variation).

Metabolomics

In order to characterize metabolic variation of (fruit) samples, nowadays efficient screening approaches are used. Such metabolomics approaches enable the parallel assessment of intensity levels of a broad range of metabolites and have great value in both phenotyping and diagnostic analyses (Fernie and Schauer 2009). Metabolomics has been carried out since the mid 1970s, but only became a standard laboratory technique in the past decade (Fernie et al. 2004). While targeted metabolite analyses focus on the analyses of specific groups of known (pepper) compounds, such as carotenoids or capsaicinoids, untargeted metabolomics approaches allow the simultaneous detection of metabolites in a biological sample, without *a priori* knowledge of the identity of the metabolites detected. Two techniques dominate metabolite profiling strategies: (i) mass spectrometry (MS) and (ii) nuclear magnetic resonance (NMR). These untargeted profiling approaches have been used to obtain an overview of the metabolic diversity in germplasm collections, such as Arabidopsis (MS; Keurentjes et al. 2006), tomato (MS; Tikunov et al. 2005) and pepper (NMR; Ritota et al. 2010).

Because these high-throughput methods generate large datasets, special analysis and visualization techniques were required to extract relevant information for elucidating the function of proteins and metabolites, the interactions between these molecules and the underlying regulatory mechanisms. Statistical analysis of such data is non-trivial since in many cases the number of metabolites outweighs the number of samples, which led to a new standard in the design of (metabol)omics

experiments: the so-called $p > n$ paradigm. Under this new paradigm, the number of independent subjects n (e.g. fruit samples) is much smaller than the number of variables p (e.g. number of metabolites in a GC-MS profile) that is analyzed. While in classical settings few pre-specified null hypotheses are evaluated, now simultaneous testing of hundreds or thousands of hypotheses needs to be performed. Classical statistical methods typically require the number of independent subjects to be large and a multiple of the number of variables in order to avoid collinearity and overfit (Dunkler et al. 2011). However, the number of samples that can be considered for a high-throughput metabolomics experiment is often limited due to the technical and economical limitations of the experiment. Hence, new statistical methods had to be developed to handle ~omics data, like Lasso (Tibshirani 1996) or Random Forest regression (Breiman 2001).

Although knowledge on the regulation of metabolite formation is increasing, for thousands of metabolites, their function in the plant, their biosynthetic pathway and the regulation thereof is still unknown. Metabolomics in combination with statistical analyses help to understand the metabolic pathways responsible for the main metabolic contrasts between genotypes and, in combination with genetic analyses, to identify QTLs and genes underlying key steps in metabolic pathways. In addition, metabolic profiling allows the detection and subsequent identification of unknown compounds correlating with a trait of interest, such as flavor. Although the costs and the extent of heritability of studied metabolites needs to be taken into account, it is beyond doubt that the concept of metabolomics-assisted breeding is a novel and powerful approach leading to new targets for breeding programs aimed at the improvement of metabolite-based quality traits (Fernie and Schauer 2009).

Genetics in pepper

To study the genetics of traits in a specific crop some prerequisites exist, such as availability of polymorphic markers, (a) genetic linkage map(s) and preferably genome sequence knowledge. The genome of *Capsicum* is diploid and consists of n =12 chromosomes with an estimated haploid genome size of 3.3–3.6 Gb (Moscone et al. 2003, Arumuganathan and Earle 1991). Similar to other members of the *Solanaceae* family such as tomato, pepper has been used as a model organism for classical and molecular genetics analyses. Over the past 25 years numerous inter- and intra-specific pepper maps have been published, of which a good overview till 2006 is given by Kumar et al. (2011). The maps have been used for the determination of syntenic relationships, gene tagging, marker-assisted selection and gene cloning. A major limitation of these maps is, however, that they have been constructed with

markers that are not high throughput or gene-based, such as AFLP, RFLP, SSR and RAPD markers. More recent therefore, a genetic linkage map based on COSII (conserved ortholog set II) markers has been published, giving insight in pepper-tomato synteny (Wu et al. 2009).

From among many types of molecular markers that have been developed during the past three decades, single nucleotide polymorphisms (SNPs) are the most attractive ones for breeding (Rafalski 2002). The majority of SNPs are bi-allelic, they are easily scored and can be tightly linked to or are the actual cause of phenotypic differences in traits. Moreover, there are several high-throughput technologies based on allele-specific PCR, hybridization and single base-pair extension which make them cost-effective for assaying large numbers of genotypes. SNPs are also extremely abundant in pepper, as next generation sequencing of the transcriptome yielded more than 22,000 high-quality putative SNPs among three pepper genotypes (Ashrafi et al. 2012) and nearly 12,000 SNPs among two other genotypes (Nicolai et al. 2012). A public whole genome sequence of pepper is not yet available, but is expected at the start of 2014 (UC Davis-Seoul National University collaboration).

Making use of aforementioned mapping populations numerous genes and QTLs have been mapped. Major attention has been given to disease resistance and color genes (reviewed by Crosby 2008) and capsaicinoid content (reviewed by Kumar et al. 2011). The recent transcriptome sequencing initiatives (Ashrafi et al. 2012, Nicolai et al. 2012) have also given a boost to mapping of the quantitative trait yield and some other physiological and fruit quality characters (Alimi et al. 2013, Yarnes et al. 2013). Genetic studies in pepper relating to both volatile and non-volatile compounds (other than capsaicinoids) contributing to flavor, however, are still lacking.

Objectives and outline of this thesis

In order to pave the way to a better understanding, prediction and improvement of pepper flavor, at the beginning of the project we formulated the following main objectives:

1. Characterization and identification of volatile and non-volatile compounds of *Capsicum* in relation to flavor and yield.

2. Study the genetics of identified (non-)volatile compounds, which contribute to flavor.

3. Introduce identified (non-)volatile compounds in pre-breeding lines for improvement of overall flavor with use of marker assisted breeding.

In this thesis therefore, we measured volatile and non-volatile compounds of mature fruits from a diverse panel of genotypes that represented, (i) roughly the flavor variation in the commercial *Capsicum annuum* breeding program of Rijk Zwaan, (ii) parents of available mapping populations (Table 1) and (iii) some genotypes that were expected to have extraordinary flavors.

Table 1. Mapping populations available at the start of the project

Type	Donor parent	Recurrent parent	Generation
RIL	*C. frutescens* BG 2814-6	*C. annuum* NuMex RNaky	F_7
BIL	*C. baccatum pendulum* PEN45	Several *C. annuum* breeding lines[1]	BC_2S_1
BIL	*C. baccatum pendulum* PEN79	*C. annuum* Vania[2]	BC_4S_2
BIL	*C. annuum* CM334	*C. annuum* Maor	BC_3S_2

[1] For description of the population see Chapter 5. [2] *C. annuum* Turrialba was used as bridge in the first cross with PEN79.

In **Chapter 2** the results of the biochemical characterization of the complete set of 35 genotypes, that were evaluated over the whole growing season (3 harvests), are discussed and some rationale for further study of the *C. baccatum pendulum* PEN45 BIL (backcross inbred lines) population (**Chapter 5**) is given. **Chapters 3 and 4** are based on a subset of 24 non-pungent genotypes that were evaluated as well for taste by a descriptive sensory panel. In **Chapter 3** the sensory and biochemical results from the first harvest are presented, while in **Chapter 4** the results were analyzed from the 3 harvests together, giving the possibility to study the effect of harvest

time and to make flavor predictions. In **Chapter 5** the results from the PEN45 BIL and derived NILs (near-isogenic lines) are discussed. Finally, the overall results and perspectives for breeding are outlined in the general discussion (**Chapter 6**).

Chapter 2

Characterization of volatile and non-volatile compounds of cultivated pepper (*Capsicum annuum*) and some wild relatives

Updated version of the publication:
P.M. Eggink, J.P.W. Haanstra, Y. Tikunov, A.G. Bovy, R.G.F. Visser
Advances in Genetics and Breeding of Capsicum and Eggplant,
Editorial de la Universitat Politècnica de València, Spain, (2010), pp. 251-259.

Abstract

In this study volatile and non-volatile compounds and several agronomical important parameters were measured in mature fruits of elite sweet pepper breeding lines and hybrids and several genebank accessions from different *Capsicum* species. The sweet pepper breeding lines and hybrids were chosen to roughly represent the expected variation in flavor of *Capsicum annuum* in the Rijk Zwaan germplasm. The genebank accessions were either chosen because they were expected to have unique combinations of aromas and flavors, according to experience and/or literature, or were parents of mapping populations. The biochemical profiling allowed visualization of between- and within-species metabolic variation and stability during the year. In general, total soluble solids content (Brix) was genotype-dependent and fluctuated only slightly throughout the growing season, with uncultivated genotypes showing the largest changes. The species *C. chinense, C. baccatum* var. *pendulum* and *C. annuum* could be clearly separated by principal components analysis based on profiles of 391 volatile compounds. Especially for breeding purposes it seems to be interesting to study this variation in more detail, trying to unravel the complex genetics of the different pepper flavor aspects.

Keywords: biochemical profiling, flavor, SPME-GC-MS, multivariate analysis, PCA.

Introduction

Flavor is an important quality parameter for fruits and vegetables. External qualities such as color, texture and shape are relatively easy to evaluate by both producers and consumers. However, evaluation of flavor attributes is more complex. In tomato flavor research measuring physical, biochemical and sensory properties, the latter were considered the most difficult to quantify (Fulton et al. 2002). Flavor of fruits and vegetables, as perceived during consumption has been defined as the overall sensation provided by the interaction of taste, odor, mouth feel, sight and sound. The composition of non-volatile compounds influences mainly the sensory perceived taste, while the aroma is affected by volatile compounds (Luning 1994b).

Although literature addressing flavor of some fruit crops, like tomato, strawberry, peach or melon, is abundant, specific research for the fruit crop pepper (*Capsicum annuum*) is limited. Pepper fruits are commonly used in the diet because of their typical color, pungency, taste and/or distinct aroma (Govindarajan 1985). Peppers are eaten fresh or processed, as unripe (green or white) or ripe (e.g. red, yellow and

orange) fruits. In the breeding of pepper, the factors production and quality (e.g. shelf life, firmness and disease resistances) are of main interest. However, since consumers have become more critical, attention in pepper, like in tomato, is shifting towards flavor as an important quality parameter (Verheul 2008).

Research on pepper flavor has mainly focused on characterization of volatile and non-volatile component variation in cultivated and/or wild species (e.g. Buttery et al. 1969, Jarret et al. 2007, Kollmannsberger et al. 2007). However, correlations between flavor components and sensory evaluations by taste or odor panels are generally missing. We aim to combine biochemical and agronomical analyses with sensory evaluations in order to elucidate the genetic and biochemical basis underlying pepper fruit flavor and, eventually, define strategies to predict and control flavor of fresh pepper. In this paper, initial results of agronomic evaluations, Brix measurements and volatile profiling will be discussed.

Materials and methods

Plant material

In this study, elite pepper breeding lines and hybrids provided by Rijk Zwaan, and several genebank accessions from multiple *Capsicum* species were used (Table 1). The pepper breeding lines and hybrids were chosen to roughly represent the variation in flavor of *C. annuum* in the germplasm of Rijk Zwaan. The genebank accessions were either chosen because they were expected to have unique combinations of aromas and flavors, according to experience and/or literature, or to be parents of available mapping populations. In 2008, the genotypes were grown in soil in a greenhouse at Rijk Zwaan (De Lier, The Netherlands), according to standard Dutch pepper management conditions with 2.5 plants/m^2. Potential shading effects, because of the diverse nature of the genotypes, were avoided by ordering the plants by (expected) plant height in the greenhouse in 3 separate blocks (i.e. tall, intermediate and short plants). All genotypes were grown in 3 plots of 5 plants, which were randomized within the separate blocks. From the beginning of May till the end of September 2008, all completely (95-100%) colored fruits were harvested, counted and weighed on a (bi)weekly base. In that period, 9 harvests, evenly spread over the season, were used for biochemical measurements, of which 3 harvests (29 May, 31 July and 4 September) were also used for sensory evaluation. After harvesting, fruits were stored in a climate room at 20°C with 80% relative humidity for 4-5 days to optimize ripening. For each individual repetition of the genotypes, a selection of 5-8 fruits was pooled to make a representative fruit sample. Fruits were cut (top

and bottom parts were discarded) in 1-2 cm pieces, mixed and seeds were removed. For fruits subjected to sensory analysis, half of the fruit pieces of each sample were immediately frozen in liquid nitrogen, ground in an electric mill and stored at -80°C while the other half was used for flavor evaluation. Fruits of harvests that were only used for biochemical measurements were prepared similarly, but were freshly processed prior to freezing in liquid nitrogen and stored at -80°C.

Metabolic profiling

The profiling of volatile metabolites was performed using headspace SPME-GC-MS, as described in Tikunov et al. 2005. Frozen fruit powder (1 g fresh weight) was weighed in a 5-ml screw-cap vial, closed and incubated at 30°C for 10 minutes. An EDTA-NaOH water solution was prepared by adjusting of 100 mM EDTA to pH of 7.5 with NaOH. Then, 1 ml of the EDTA-NaOH solution was added to the sample to a final EDTA concentration of 50 mM. Solid CaCl2 was then immediately added to give a final concentration of 5 M. The closed vials were then sonicated for 5 minutes. A 1 ml aliquot of the pulp was transferred into a 10-ml crimp cap vial (Waters), capped and used for SPME-GC-MS analysis.

Volatiles were automatically extracted from headspace and injected into the GC-MS via a Combi PAL autosampler (CTC Analytics AG). Headspace volatiles were extracted by exposing a 65 µm PDMS-DVB SPME fiber (Supelco) to the vial headspace for 20 minutes under continuous agitation and heating at 50°C. The fiber was inserted into a GC 8000 (Fisons Instruments) injection port and volatiles were desorbed for 1 min at 250°C. Chromatography was performed on an HP-5 (50m X 0.32 mm X 1.05 µm) column with helium as carrier gas (37 kPa). The GC interface and MS source temperatures were 260°C and 250°C, respectively. The GC temperature program began at 45°C (2 min), was then raised to 250°C at a rate of 5°C/min and finally held at 250°C for 5 min. The total run time including oven cooling was 60 min. Mass spectra in the 35-400 *m/z* range were recorded by an MD800 electron impact MS (Fisons Instruments) at a scanning speed of 2.8 scans/sec and an ionization energy of 70 eV. The chromatography and spectral data were evaluated using "XcaliburTM" software (http://www.thermo.com).

Clear supernatants of shortly centrifuged samples were used for refractive index measurement of total soluble solids content (TSS; °Brix) and for an enzymatic determination of glucose, fructose and sucrose (Velterop and Vos 2001). Anion exchange chromatography on the same supernatants was used for citric, malic and ascorbic acid determination based on standard protocols (Dionex Corporation, Sunnyvale, CA; http://www.dionex.com/ Application Note 143 "Determination

of Organic Acids in Fruit Juices"). Dry matter content was calculated by drying weighed samples at 60-80°C for up to 48h in a standard oven.

GC-MS data processing

The GC–MS profiles derived using the SPME-GC–MS method were processed by the MetAlignTM software package (http://www.metalign.nl) for baseline correction, noise estimation and ion-wise mass spectral alignment. The Multivariate Mass Spectral Reconstruction (MMSR) approach (Tikunov et al. 2005) was used to reduce data to volatile compound mass spectra. Each compound was represented by a single selective ion fragment in the following multivariate data analysis. The compounds (number of fragment ions in a mass spectrum ≥5) were then subjected to a tentative identification using the NIST mass spectral library (http://www.nist.gov). Identities were assigned to compounds with a forward match factor (fmf) ≥ 700. The rest of the compounds were considered of unknown identity. Identities of 21 volatiles were confirmed by authentic chemical standards.

Volatile data analysis

The (non-)volatile data has been analyzed using GeneMaths XT version 2.0 (http://www.applied-maths.com). The data sets have been log2 transformed and normalized to the mean. Principal components analysis (PCA) implemented in GeneMaths was used for unsupervised cluster analysis of the metabolites. Pearson's correlation coefficient was used as a measure for metabolite-metabolite correlation and hierarchical clustering.

Results and discussion

Agronomical evaluations

In correspondence to the genetic diversity in our collection of 35 *Capsicum* genotypes representing 4 different species, we found a wide range of agronomical characteristics (Table 1). Fruits were ranging 0.5-22 cm in length and 0.5-8.5 cm in width, within the fruit types blocky, dulce italiano, dolma, kapya, lamuyo, conical, elongated, round and Habenero. The majority of the genotypes were red, as this is the predominant color in cultivated and wild material; yellow and orange genotypes were less represented. The accession Chinense-WA segregated for yellow and red fruits. Therefore biochemical measurements and sensory evaluations of this accession were

performed on samples of the separate fruit colors. Due to the fact that we were also interested in studying some non-cultivated accessions, several pungent genotypes were included in the analyses. Total yield was reported throughout the complete analysis period (May-September) and large differences were observed (0.1-15.3 kg/m^2). Since *C. frutescens* BG2814-6 and *C. annuum* Turrialba yield a large amount of very small fruits, the total yield was estimated based on the approximate amount of harvested fruits and the average fruit weight.

Biochemical analyses

Table 1. Description of *Capsicum* genotypes evaluated for fruit quality attributes.

Genotype	Species	Fruit type	Size[1] (cm)	Color	Pungency	°Brix[2]	Yield[3] (kg/m^2)
Mazurka[4]	*C. annuum* (elite)	Blocky	8 x 8	Red	Sweet	7.6	12.1
Hybrid 1 */**	*C. annuum* (elite)	Blocky	8 x 8	Red	Sweet	8.0	12.8
Line A	*C. annuum* (elite)	Blocky	8 x 8	Red	Sweet	7.6	11.7
Line B	*C. annuum* (elite)	Blocky	8.5 x 8	Red	Sweet	7.8	9.6
Line C *	*C. annuum* (elite)	Blocky	8.5 x 8	Red	Sweet	8.4	12.9
Line D *	*C. annuum* (elite)	Blocky	8 x 8	Red	Sweet	7.8	9.1
Line F *	*C. annuum* (elite)	Blocky	9 x 8	Yellow	Sweet	5.6	11.8
Line G **	*C. annuum* (elite)	Blocky	8 x 8	Yellow	Sweet	6.4	15.3
Line H	*C. annuum* (elite)	Blocky	8 x 8	Yellow	Sweet	8.0	12.9
Line I	*C. annuum* (elite)	Blocky	8 x 8.5	Yellow	Sweet	7.2	14.6
Line J **	*C. annuum* (elite)	Blocky	8 x 8.5	Orange	Sweet	7.4	12.4
Line K	*C. annuum* (elite)	Mini block	5 x 5	Orange	Sweet	8.3	7.2
Hybrid 2 *	*C. annuum* (elite)	Dulce italiano	20 x 4	Red	Sweet	9.4	10.9
Hybrid 3	*C. annuum* (elite)	Dulce italiano	22 x 4.5	Red	Sweet	9.5	13.0
Line L */**	*C. annuum* (elite)	Dulce italiano	22 x 4	Red	Sweet	7.7	11.5
Line M *	*C. annuum* (elite)	Dulce italiano	18 x 4.5	Red	Sweet	9.4	9.8
Line O	*C. annuum* (elite)	Dulce italiano	22 x 4	Red	Sweet	7.6	11.6
Line P	*C. annuum* (elite)	Dulce italiano	22 x 4	Red	Pungent	7.8	13.5
Line E	*C. annuum* (elite)	Dolma	7 x 6.5	Red	Sweet	8.7	8.7
Line N	*C. annuum* (elite)	Kapya	12 x 4	Red	Sweet	8.3	8.8
Piquillo **	*C. annuum*	Conical	9 x 4	Red	Sweet	10.5	6.6
Buran	*C. annuum*	Lamuyo	10 x 7	Red	Sweet	9.1	11.0
PBC1405 */**	*C. annuum*[5]	Elongated	18 x 2	Red	Sweet	9.8	8.8
PI543188	*C. annuum*[6]	Conical	10 x 4	Red	Pungent	7.8	5.9
Antillais Caribbean */**	*C. chinense*	Habanero	5 x 5	Red	Pungent	6.3	6.4
Chinense-WA */**	*C. chinense*	Habanero	5 x 5	Red/ yellow[7]	Pungent	6.0	8.3

Genotype	Species	Fruit type	Size[1] (cm)	Color	Pungency	°Brix[2]	Yield[3] (kg/m²)
BG 2814-6 **	*C. frutescens*	Round	0.5-1	Red	Pungent	25.7	0.1[8]
Numex RNaky	*C. annuum*	Dulce Italiano	20 x 4	Red	Pungent	9.2	9.9
PEN-45 */**	*C. baccatum pendulum*	Conical	6-7 x 2	Red	Pungent	11.2	8.9
PEN-79 *	*C. baccatum pendulum*	Conical	6-7 x 2	Red	Pungent	11.8	11.1
Turrialba **	*C. annuum*	Round	1-1.5	Red	Pungent	14.8	0.5[8]
Vania	*C. annuum*	Lamuyo	14 x 8	Red	Sweet	9.0	10.1
CM334	*C. annuum*	Conical	6-7 x 4	Red	Pungent	9.0	2.4
Maor	*C. annuum*	Blocky	8 x 8	Red	Sweet	7.9	12.0
Perennial **	*C. annuum*	Elongated	3-4 x 1	Red	Pungent	13.0	1.5

[1] Size is indicated by length x width, [2]Average total soluble solids of the fruit samples (9/genotype) that were used for sensory evaluation, [3] Average yield in the harvesting period May through September, [4] Control variety (e.g. Luning et al. 1994a and 1994b), [5] Accession formerly classified as *C. baccatum* (AVRDC), [6] Accession formerly classified as *C. chinense* (USDA), [7] Accession is segregating for yellow and red fruits, [8] Yield is estimated based on the approximate amount of harvested fruits and the average fruit weight *Subset containing genotypes most contrasting in (non-)volatiles and flavour, **Genotypes included in bulk reference sample.

The non-volatile compounds including total soluble solids, sugars (fructose, glucose and sucrose) and acids (malic, citric and ascorbic acid) completed with dry matter were measured on 9 harvests in the period May-September, evenly spread over the season, of which 3 harvests (29 May, 31 July and 4 September) were also used for sensory evaluation and volatile profiling (see Chapter 3 and 4). Sucrose concentrations turned out to be under the detection limit (0.3 g/100 g fresh weight) of our enzymatic determination method. Evaluations of the other non-volatiles are reported in Table 2 from the harvest of 29 May. All 35 genotypes of the mentioned 3 harvests were included in the (non-)volatile measurement, whereas from the other 6 harvests only a subset (marked * in Table 1) of 12 genotypes, which were most contrasting in flavor and (non-)volatile profile, were included in non-volatile analyses. Figure 1 shows an overview of the total soluble solids content (TSS) of the 12 genotypes in the subset, which gives an impression of the non-volatile compound concentration behavior during the year and the variability between repetitions of the same genotype in the experiment. In general TSS fluctuated only slightly throughout the growing season with relatively small standard errors of the means, indicating uniformity of the experimental setup. Uncultivated genotypes, like *C.baccatum* var. *pendulum* PEN45 and PEN79 showed the largest fluctuations, mainly at the start of the experiment. In addition, to the stability of TSS compounds, their effect on yield is also an important

breeding parameter. In the total set of 35 genotypes, the correlation between TSS and yield is -0.64 (41.4% explained variance), whereas in elite material this correlation is -0.38 (14.3% explained variance). The negative relationship between TSS and yield has been observed before. Utilizing 20 years of processing tomato field data, Grandillo and co-workers (1999) reported a negative correlation between °Brix and yield ranging between -0.23 and -0.57 depending on period and environment.

Using SPME-GC-MS 391 volatile compounds were detected, of which 189 compounds were of unknown origin [fmf<700]). This number of pepper volatiles is in the same order of magnitude as the number of compounds (322) found in a diverse set of tomato genotypes (Tikunov et al. 2005). In Figure 2, the hierarchical clustering from 16 genotypes (harvest 29 May) based on intensity patterns of all measured volatile compounds is shown. Genotype repetitions and bulk samples, consisting of a balanced mixture of 12 representative genotypes (marked ** in Table 1), were used as reference in all SPME-GC-MS measurements, and generally clustered together confirming the quality of the data. Principal components analysis (PCA) proved to be a powerful method to visualize differences between the genotypes, discriminating between- and within-species variation. The species *C. chinense*, *C. baccatum* var. *pendulum* and *C. annuum* clustered separately along the primary axis (58.3% explained variance). A group of at least 15 saturated and unsaturated esters were mainly responsible for the individual grouping of the *C. chinense* accessions (data not shown), which is in line with the findings of Rodriguez-Burruezo and colleagues (2010). Separation along the vertical axis (8.9% explained variance) is mainly based on within-species variation.

Conclusion and continuation

The biochemical profiling allowed visualization of between- and within-species (non-)volatile variation and stability during the year. The PCA plot (Fig. 2b) shows individual grouping of *C. chinense, C. baccatum* var. *pendulum* and *C. annuum*, indicating potentially interesting volatile variation present in the former two groups. In addition, the variation within the *C.annuum* (elite) group itself gives sufficient reason to justify more detailed study. In both cases, mapping populations resulting from crossing extreme and contrasting genotypes would possibly allow the unraveling of the different aspects of pepper flavor genetics. Because of the complex nature of flavor, thorough biochemical, sensory and agronomical evaluation in combination with QTL mapping will be needed.

Finally, in addition to biochemical profiling, the genotypes have been subjected to sensory evaluation by a trained descriptive expert panel (data not shown). In

chapter 3 and 4 we described the relation between specific biochemical compounds and sensory attributes.

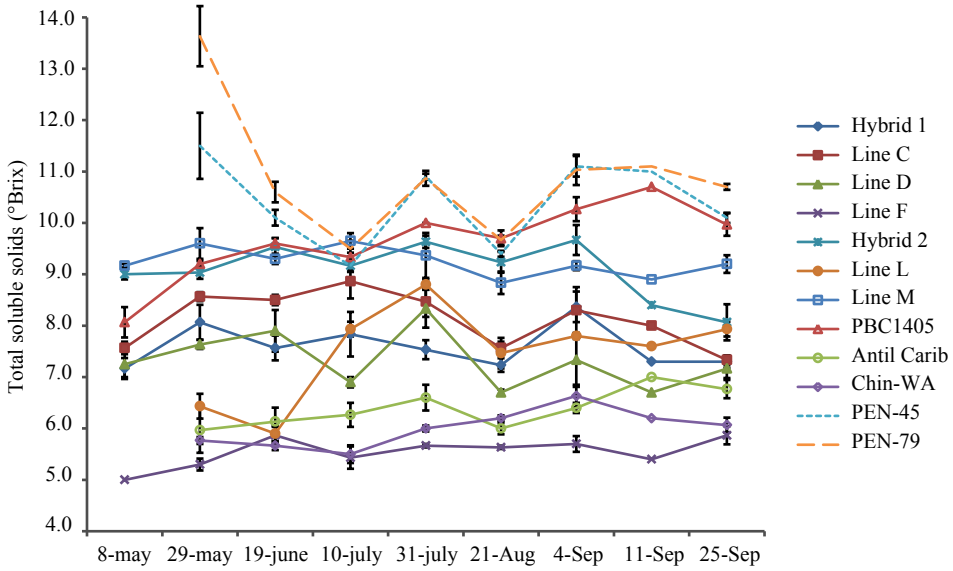


Figure 1. Total soluble solids content of the 12 genotypes in the subset during the year. Mean values and standard errors from three measurements (error bars) are shown.

Table 2. Evaluations of dry matter, total soluble solids, sugars and acids

Genotype[1]	DM [2] %	°Brix	Glucose[3]	Fructose[3]	Malic acid[4]	Citric acid[4]	Ascorbic acid[4]
Mazurka	9.15	7.67	2.83	2.81	16.63	420.5	174.4
Hybrid 1	9.35	8.07	3.11	3.05	18.08	374.9	157.4
Line A	9.14	7.63	2.97	2.71	19.63	396.6	173.3
Line B	9.35	7.90	3.02	3.16	14.96	372.2	165.8
Line C	10.14	8.57	3.14	3.34	20.66	382.7	200.4
Line D	8.95	7.63	2.77	2.80	16.46	388.2	139.7
Line F	6.11	5.30	1.95	1.89	11.71	185.7	137.9
Line G	6.77	5.53	1.79	1.95	15.40	252.0	146.2
Line H	9.03	7.83	2.87	3.06	22.90	367.8	169.7
Line I	7.96	6.90	2.26	2.51	19.42	315.6	178.4
Line J	8.58	7.27	2.72	2.74	14.97	338.4	201.4
Line K	9.78	8.63	3.25	3.11	76.05	325.6	174.9
Hybrid 2	10.46	9.03	3.29	3.51	27.57	381.0	181.5
Hybrid 3	11.41	9.90	3.81	3.75	25.51	431.3	209.9

Genotype[1]	DM [2] %	°Brix	Glucose[3]	Fructose[3]	Malic acid[4]	Citric acid[4]	Ascorbic acid[4]
Line L	8.08	6.43	2.07	2.09	32.74	286.2	195.2
Line M	10.91	9.60	3.51	3.74	29.07	334.5	186.5
Line O	8.12	6.37	2.23	2.15	31.65	301.4	190.1
Line P	9.80	6.63	2.34	2.37	27.19	304.9	200.7
Line E	11.19	9.03	3.17	3.29	30.19	450.7	211.0
Line N	10.97	8.97	3.23	3.27	159.27	275.9	188.2
Piquillo	13.65	10.37	3.39	3.61	56.94	542.3	215.7
Buran	11.04	9.43	3.49	3.39	30.44	363.2	195.0
PBC1405	12.56	9.20	3.04	3.11	39.46	609.7	247.1
PI 543188	9.90	8.20	2.47	2.68	34.44	496.4	175.5
Antillais Caribbean	9.60	5.97	1.56	1.83	74.03	328.7	94.2
Chinense-WA red	8.40	5.53	1.62	1.51	238.09	144.8	101.1
Chinense-WA yellow	8.80	5.77	1.50	1.49	246.68	131.2	90.8
BG 2814-6	42.30	25.70	1.98	1.89	610.61	3962.1	188.3
NuMex RNaky	13.10	9.27	2.99	2.95	29.49	538.9	290.4
PEN-45	14.10	11.50	3.93	3.24	126.65	1054.2	55.6
PEN-79	14.90	13.63	4.14	3.36	192.98	1312.2	53.6
Turrialba	37.20	14.50	0.96	0.69	849.95	2110.4	112.4
Vania	10.21	8.80	3.47	3.36	35.05	318.7	194.6
CM334	13.60	10.03	2.83	3.45	49.35	634.1	217.2
Maor	9.57	8.27	3.21	3.06	27.04	359.0	170.8
Perennial	21.90	na	2.08	1.88	238.47	1752.9	388.5

[1]Values are averages of 3 individual biological replicates from the harvest of 29 May. [2] Dry matter, [3] g/100 g fresh weight, [4] mg/100 g fresh weight.

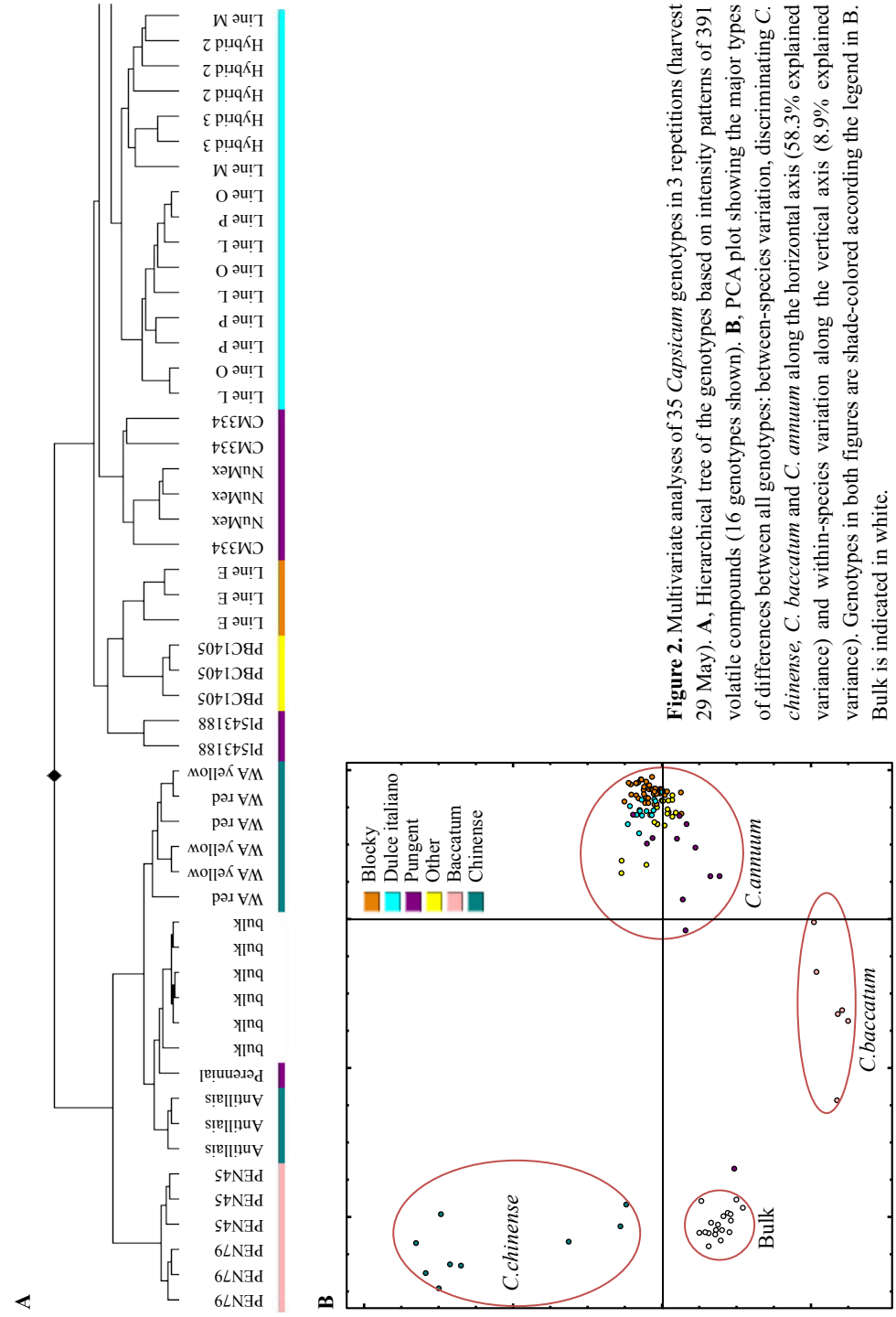


Figure 2. Multivariate analyses of 35 *Capsicum* genotypes in 3 repetitions (harvest 29 May). **A**, Hierarchical tree of the genotypes based on intensity patterns of 391 volatile compounds (16 genotypes shown). **B**, PCA plot showing the major types of differences between all genotypes: between-species variation, discriminating *C. chinense*, *C. baccatum* and *C. annuum* along the horizontal axis (58.3% explained variance) and within-species variation along the vertical axis (8.9% explained variance). Genotypes in both figures are shade-colored according the legend in B. Bulk is indicated in white.

Acknowledgements

The authors kindly acknowledge Harry Jonker and Yvonne Birnbaum at Plant Research International for performing the SPME-GC-MS analyses. In addition we thank Sjaak van Heusden for technical support and useful discussions. Finally we are grateful to Laure Flament, Suzanne de Wit, Femke Willeboordse, Gerald Freymark, Sander Bos, Tineke Benning and Paula de Grauw for performing a massive job on sample preparation and non-volatile measurements, and all other people at Rijk Zwaan who took care of perfect greenhouse management.

Chapter 3

A taste of sweet pepper: Volatile and non-volatile chemical composition of fresh sweet pepper (*Capsicum annuum*) in relation to sensory evaluation of taste

P.M. Eggink, C. Maliepaard, Y. Tikunov, J.P.W. Haanstra,
A.G. Bovy, R.G.F. Visser
Food Chemistry (2012) 132: 301-310

Abstract

In this study volatile and non-volatile compounds as well as some breeding parameters were measured in mature fruits of elite sweet pepper (*Capsicum annuum*) lines and hybrids from a commercial breeding program, several cultivated genotypes and one gene bank accession. In addition, all genotypes were evaluated for taste by a trained descriptive sensory expert panel. Metabolic contrasts between genotypes were caused by clusters of volatile and non-volatile compounds, which could be related to metabolic pathways and common biochemical precursors. Clusters of phenolic derivatives, higher alkanes, sesquiterpenes and lipid derived volatiles formed the major determinants of the genotypic differences. Flavor was described with use of 14 taste attributes of which the texture related attributes and the sweet-sour contrast were the most discriminatory factors. The attributes juiciness, toughness, crunchiness, stickiness, sweetness, aroma, sourness and fruit/apple taste could be significantly predicted with combined volatile and non-volatile data. Fructose and (E)-2-hexen-1-ol were highly correlated with aroma, fruity/apple taste and sweetness. New relations were found for fruity/apple taste and sweetness with the compounds p-menth-1-en-9-al, (E)-β-ocimene, (Z)-2-penten-1-ol and (E)-geranylacetone. Based on overall biochemical and sensory results, the perspectives for flavor improvement by breeding are discussed.

Keywords: sensory evaluation; biochemical profiling; metabolomics; SPME-GC-MS; multivariate analysis; Random Forest.

Introduction

Sweet and hot peppers (*Capsicum annuum*) are cultivated worldwide and form important food ingredients. Fruits are commonly used in diets because of their typical color, pungency, taste and/or distinct aroma (Govindarajan 1985). Especially, sweet peppers are eaten fresh or processed, as unripe (green or white) or ripe (e.g. red, yellow and orange) fruits. In the breeding process of pepper the factors production (yield) and quality, e.g. shelf life, firmness and disease resistance, are of main interest. Consumers have, however, become more critical in the last decade, resulting in a need towards flavor as a more important quality parameter in pepper breeding (Verheul 2008). Flavor of fruits and vegetables, as perceived during consumption has been defined as the overall sensation provided by the interaction of taste, aroma, mouth feel, sight and sound (Luning et al. 1994b). Especially the interplay among all these parameters in combination with different consumer preferences make flavor

such a difficult property to quantify in an objective way. Literature about flavor of some fruit crops, like tomato, strawberry, kiwi or melon is abundant, however, specific research addressing sweet pepper flavor is limited. Studies so far mainly focused on characterization of variation for volatile and/or non-volatile components in cultivated or wild species (e.g. Buttery et al. 1969, Jarret et al. 2007, Rodriguez-Burruezo et al. 2010). Correlations between biochemical compounds and sensory attributes scored by taste panels are generally missing.

In the present study we characterized flavor in a broad germplasm panel from a commercial breeding program completed with few cultivated genotypes and a gene bank accession. Flavor was objectively quantified by thorough biochemical profiling in combination with sensory evaluations by a trained expert panel. The overall results make it possible to link individual taste attributes to volatile and non-volatile compounds. Taking their effect on total yield into account, our results form a starting point towards directed flavor breeding in pepper.

Materials and methods

Plant material

In this study 24 non-pungent *Capsicum annuum* accessions from the broader collection of *Capsicum* genotypes described in Eggink et al. (2010) were used. The panel consisted of elite pepper breeding lines and hybrids (provided by Rijk Zwaan), several cultivated genotypes (landraces or old hybrids) and one gene bank accession (Table 1). The pepper breeding lines and hybrids were chosen to roughly represent the flavor variation in the *C. annuum* germplasm of a typical commercial breeding program. The cultivated variety Piquillo was chosen as it is famous in the Mediterranean region for its full taste and rich aroma. Cultivar Buran was reported to be a very sweet lamuyo type (http://www.seedsavers.org) and cv. Vania and cv. Maor have been used as parents in publicly available mapping populations (e.g. Lefebvre et al. 2003, Ben Chaim et al. 2003). Finally, PBC1405 was included, because according to the AVRDC gene bank (http://www.avrdc.org), it is a non-pungent *C. baccatum* accession, which is rather unique as most accessions from wild *Capsicum* species are pungent.

In 2008, the genotypes were grown in soil in a greenhouse at Rijk Zwaan (De Lier, The Netherlands), according to standard Dutch pepper management conditions with 2.5 plants/m^2. Potential shading effects, because of the diverse nature of the genotypes, were avoided by ordering the plants by (expected) plant height in the greenhouse in 3 separate blocks (i.e. tall, intermediate and short plants). All genotypes were grown

in 3 plots of 5 plants, which were randomized within the separate blocks. From the beginning of May till the end of September 2008, all ripe (95-100% colored) fruits were harvested, counted and weighed on a (bi)weekly base. The harvest of May 29 was used for biochemical measurements and sensory evaluation. After harvesting, fruits were stored in a climate room at 20°C with 80% relative humidity for 4-5 days to optimize ripening. This is standard procedure to mimic the Dutch commercial system. From each individual repetition of the genotypes, a selection of 5-8 fruits was pooled to make a representative fruit sample. Fruits were cut (top and bottom parts were discarded) in 1-2 cm pieces, mixed and seeds were removed. Half of the fruit pieces from each sample were immediately frozen in liquid nitrogen, ground in an electric mill and stored at -80°C while the other half was used for flavor evaluation.

Sensory analysis

The 24 genotypes were subjected to sensory evaluation by a trained descriptive expert panel at Wageningen UR Greenhouse Horticulture (WUR-GH, Bleiswijk, The Netherlands). The sweet genotypes were evaluated by, on average, 16 taste panelists in a randomized setup, split over 2 subsequent days. On both days, 2 sessions with 6 genotypes and a reference (commercial red blocky *C. annuum* hybrid) were evaluated per panelist. Each panelist received 5 fruit pieces per sample. The setup was chosen in such a way that each panelist evaluated one repetition of all genotypes and that each sample was evaluated by 5-6 panelists. The expert panel evaluated 14 attributes on a scale from 0 to 100 to describe the taste sensation in the mouth/throat which were: crunchiness, stickiness of the skin, toughness, juiciness, sweetness, sourness, aroma intensity, grassiness, green bean taste, carrot taste, fruity/apple taste, perfume taste, petrochemical taste and musty taste.

Metabolic profiling

The profiling of volatile metabolites was performed using headspace SPME-GC-MS, as described by Tikunov and colleagues (2005). Frozen fruit powder (1 g) was weighed in a 5-ml screw-cap vial, closed and incubated at 30°C for 10 minutes. An EDTA-NaOH water solution was prepared by adjusting of 100 mM EDTA to pH of 7.5 with NaOH. Then, 1 ml of the EDTA-NaOH solution was added to the sample to a final EDTA concentration of 50 mM. Solid CaCl2 was then immediately added to give a final concentration of 5 M. The closed vials were then sonicated for 5 minutes. A 1 ml aliquot of this solution was transferred into a 10-ml crimp cap vial (Waters), capped and used for SPME-GC-MS analysis.

Table 1. Description of Capsicum annuum genotypes evaluated for fruit quality attributes

Genotype	Origin	Source country	Fruit type	Size[1] (cm)	Color	Yield[2] (kg/m^2)
Mazurka[3]	Elite	Netherlands	Blocky	8 x 8	Red	12.1 ± 0.1
Hybrid 1	Elite	Netherlands	Blocky	8 x 8	Red	12.8 ± 0.8
Line A	Elite	Netherlands	Blocky	8 x 8	Red	11.7 ± 0.3
Line B	Elite	Netherlands	Blocky	8.5 x 8	Red	9.6 ± 0.9
Line C	Elite	Netherlands	Blocky	8.5 x 8	Red	12.9 ± 0.3
Line D	Elite	Netherlands	Blocky	8 x 8	Red	9.1 ± 0.7
Line F	Elite	Netherlands	Blocky	9 x 8	Yellow	11.8 ± 1.3
Line G	Elite	Netherlands	Blocky	8 x 8	Yellow	15.3 ± 0.6
Line H	Elite	Netherlands	Blocky	8 x 8	Yellow	12.9 ± 0.5
Line I	Elite	Netherlands	Blocky	8 x 8.5	Yellow	14.6 ± 1.4
Line J	Elite	Netherlands	Blocky	8 x 8.5	Orange	12.4 ± 2.4
Line K	Elite	Netherlands	Mini block	5 x 5	Orange	7.2 ± 1.1
Hybrid 2	Elite	Italy	Dulce italiano	20 x 4	Red	10.9 ± 2.0
Hybrid 3	Elite	Italy	Dulce italiano	22 x 4.5	Red	13.0 ± 0.8
Line L	Elite	Italy	Dulce italiano	22 x 4	Red	11.5 ± 0.5
Line M	Elite	Italy	Dulce italiano	18 x 4.5	Red	9.8 ± 1.4
Line O	Elite	Italy	Dulce italiano	22 x 4	Red	11.6 ± 3.0
Line E	Elite	Turkey	Dolma	7 x 6.5	Red	8.7 ± 1.1
Line N	Elite	Turkey	Kapya	12 x 4	Red	8.8 ± 1.1
Piquillo	Cultivated	Spain	Conical	9 x 4	Red	6.6 ± 0.7
Buran	Cultivated	Poland	Lamuyo	10 x 7	Red	11.0 ± 0.3
PBC1405	Gene bank	AVRDC, Taiwan	Elongated	18 x 2	Red	8.8 ± 1.0
Vania	Cultivated	France	Lamuyo	14 x 8	Red	10.1 ± 0.9
Maor	Cultivated	Spain	Blocky	8 x 8	Red	12.0 ± 0.8

[1] Size is indicated by length x width, [2] Average yield and standard deviation in the harvesting period May through September, [3] Standard variety (reference to e.g. Luning et al. 1994a and 1994b).

Volatiles were automatically extracted from headspace and injected into the GC-MS via a Combi PAL autosampler (CTC Analytics AG). Headspace volatiles were extracted by exposing a 65 µm PDMS-DVB SPME fiber (Supelco) to the vial headspace for 20 minutes under continuous agitation and heating at 50°C. The fiber was inserted into a GC 8000 (Fisons Instruments) injection port and volatiles were desorbed for 1 min at 250°C. Chromatography was performed on an HP-5 (50m X 0.32 mm X 1.05 µm) column with helium as carrier gas (37 kPa). The GC

interface and MS source temperatures were 260°C and 250°C, respectively. The GC temperature program began at 45°C (2 min), was then raised to 250°C at a rate of 5°C/min and finally held at 250°C for 5 min. The total run time including oven cooling was 60 min. Mass spectra in the 35-400 *m/z* range were recorded by an MD800 electron impact MS (Fisons Instruments) at a scanning speed of 2.8 scans/sec and an ionization energy of 70 eV. The chromatography and spectral data were evaluated using "XcaliburTM" software (http://www.thermo.com).
Clear supernatants of shortly centrifuged samples were used for refractive index measurement of total soluble solids content (TSS; °Brix) and for an enzymatic determination of glucose, fructose and sucrose (Velterop and Vos 2001). Anion exchange chromatography on the same supernatants was used for citric, malic and ascorbic acid determination based on standard protocols (Dionex Corporation, Sunnyvale, CA; http://www.dionex.com/ Application Note 143 "Determination of Organic Acids in Fruit Juices"). Dry matter content was calculated by drying weighed samples at 60-80°C for up to 48h in a standard oven.

GC-MS data processing

The GC–MS profiles derived using the SPME-GC–MS method were processed by the MetAlignTM software package (http://www.metalign.nl) for baseline correction, noise estimation and ion-wise mass spectral alignment. The Multivariate Mass Spectral Reconstruction (MMSR) approach (Tikunov et al. 2005) was used to reduce data to volatile compound mass spectra. Each compound was represented by a single selective ion fragment in the following multivariate data analysis. The compounds (number of fragment ions in a mass spectrum ≥5) were then subjected to a tentative identification using the NIST mass spectral library (http://www.nist.gov). Identities were assigned to compounds with a forward match factor (fmf) ≥ 700 and an identity probability rank ≥ 2 (Mihaleva et al. 2009). Identities of 21 volatiles were confirmed by authentic chemical standards.

Data analysis

The sensory data was analysed in Genstat version 12 (http://www.vsni.co.uk) using a linear mixed model REML (residual maximum likelihood) analysis with genotype and replicate and their interaction as fixed terms. Sessions (tasting sessions) within replicate/genotype combinations and panelists within sessions were taken as random terms. Mean values were calculated per genotype per replicate after a correction for session and panelist effects and removal of strong outliers (if the absolute value of a

standardized residuals was larger than three residual standard deviations).

Principal components analysis (PCA) biplot visualization as implemented in GeneMaths XT version 2.0 (http://www.applied-maths.com) was used for showing relationships between and among metabolites and attributes. For these analyses the metabolite data sets were log transformed and mean centered. Pearson's correlation coefficient was used as a measure for metabolite-metabolite correlation and for hierarchical clustering analysis (HCA) using the UPGMA algorithm.

Variance components of the sensory attributes were estimated using a variance components model with REML in Genstat version 12 with the terms genotype, replicate, their interaction, and sessions within genotype/replicate combinations and taster within sessions.

A Random Forest (Breiman 2001) regression approach was used to relate each sensory attribute (response) to the volatile and non-volatile data (predictors) and to determine importance of the individual volatiles and non-volatiles. A double ten-fold cross-validation approach was used to optimize the number of variables for each decision rule in the random forest (the 'mtry' variable in the R function to perform Random Forest) and to estimate the mean square error (on independent test samples). The performance of the models is expressed by the prediction R^2, which is calculated from the out-of-bag samples (Breiman 2001).This R^2 value therefore is not a goodness-of-fit of the data at hand but an estimate of predictive accuracy on independent (left-out) samples. Variable importance was estimated by the increase in mean square error (MSE) after permutation (Breiman 2001). Significance of the prediction R^2 and variable importance was determined by another permutation test, in which the attributes were permuted over the genotypes while retaining the original metabolic values of the genotypes. In this permuted situation, 100 new Random Forest models were constructed with calculation of the prediction R^2 and increase in MSE to estimate variable importance. Significance thresholds were determined at $P<0.05$.

Results

Genotypes

In correspondence to the genetic diversity in our collection of 24 non-pungent *Capsicum annuum* accessions, a high degree of variation was found for fruit size and yield. Fruits ranged from 5-22 cm in length and 2-8.5 cm in width, within the fruit types blocky, dulce italiano, dolma, kapya, lamuyo, conical and elongated (Table 1). The majority of the genotypes were red, as this is the predominant color in

cultivated material; yellow and orange genotypes were less represented. Total yield was measured throughout the complete growth period (May through September) and ranged from 6.6-15.3 kg/m^2. Based on (flower) morphology, the plants grown from our PBC1405 seed lot turned out to belong to *C. annuum*, while it was reported by AVRDC to be a non-pungent *C. baccatum* accession.

Identification of metabolites

The Multivariate Mass Spectral Reconstruction (MMSR) approach (Tikunov et al. 2005) was used to reduce GC-MS data to volatile compound mass spectra. In the set of 24 *C. annuum* genotypes in total 224 molecular fragment clusters were obtained, putatively representing the mass spectra of 224 individual volatile compounds. All compounds were subjected to a putative identification by matching their mass spectra to the NIST library and reliable identities (mass spectra match factor ≥ 700 and identity probability rank ≥ 2) could be assigned to 100 of them. Relative intensity patterns of all 224 compounds are given in Supplementary Table 1. In addition to the GC-MS measurements, the concentration of non-volatile flavour compounds, such as sugars (fructose, glucose and sucrose) and acids (malic, citric and ascorbic acid) was measured, completed by dry matter content and total soluble solids (TSS) determination (Table 2). Sucrose concentrations turned out to be under the detection limit (0.3 g/100 g fresh weight) of our enzymatic determination method.

Volatiles are correlated according to metabolic pathway

Hierarchical cluster analysis (HCA) was performed using intensity patterns of all 224 volatiles, concentrations of the 5 primary metabolites (*i.e.* sugars and acids) plus TSS and dry matter measurements. Few clusters of highly correlated compounds were found, which are shown as dark colored blocks in the correlation matrix (Fig. 1). NIST library matching results indicated that nine of these blocks contained compounds that have a common biochemical precursor or belong to the same metabolic pathway: phenolic derivatives (a, i), higher alkanes (c), sesquiterpenes (d), lipid derivatives (e), terpenoids (f, h) and saturated acid derivatives (g). In our 24 genotypes also the majority of primary metabolites as well as TSS and dry matter content, clustered together. Specifically, the sugars glucose and fructose grouped with citric and ascorbic acid, TSS, 3-methyl-butanoic acid, 3-methyl-3-butenylester and several compounds of unknown identity (cluster b, Fig. 1). Malic acid did not cluster with the other primary metabolites.

Table 2. Evaluations of dry matter, total soluble solids, sugars and acids

Genotype	Dry matter (%)[1]	°Brix	Glucose (g/100 g fw[2])	Fructose (g/100 g fw)	Malic acid (mg/100 g fw)	Citric acid (mg/100 g fw)	Ascorbic acid (mg/100 g fw)
Mazurka	9.15	7.67	2.83	2.81	16.63	420.5	174.4
Hybrid 1	9.35	8.07	3.11	3.05	18.08	374.9	157.4
Line A	9.14	7.63	2.97	2.71	19.63	396.6	173.3
Line B	9.35	7.90	3.02	3.16	14.96	372.2	165.8
Line C	10.14	8.57	3.14	3.34	20.66	382.7	200.4
Line D	8.95	7.63	2.77	2.80	16.46	388.2	139.7
Line F	6.11	5.30	1.95	1.89	11.71	185.7	137.9
Line G	6.77	5.53	1.79	1.95	15.40	252.0	146.2
Line H	9.03	7.83	2.87	3.06	22.90	367.8	169.7
Line I	7.96	6.90	2.26	2.51	19.42	315.6	178.4
Line J	8.58	7.27	2.72	2.74	14.97	338.4	201.4
Line K	9.78	8.63	3.25	3.11	76.05	325.6	174.9
Hybrid 2	10.46	9.03	3.29	3.51	27.57	381.0	181.5
Hybrid 3	11.41	9.90	3.81	3.75	25.51	431.3	209.9
Line L	8.08	6.43	2.07	2.09	32.74	286.2	195.2
Line M	10.91	9.60	3.51	3.74	29.07	334.5	186.5
Line O	8.12	6.37	2.23	2.15	31.65	301.4	190.1
Line E	11.19	9.03	3.17	3.29	30.19	450.7	211.0
Line N	10.97	8.97	3.23	3.27	159.27	275.9	188.2
Piquillo	13.65	10.37	3.39	3.61	56.94	542.3	215.7
Buran	11.04	9.43	3.49	3.39	30.44	363.2	195.0
PBC1405	12.56	9.20	3.04	3.11	39.46	609.7	247.1
Vania	10.21	8.80	3.47	3.36	35.05	318.7	194.6
Maor	9.57	8.27	3.21	3.06	27.04	359.0	170.8

[1]Values are averages of 3 individual biological replicates. Individual measurements are included in Supplementary Table 1. [2] Fresh Weight.

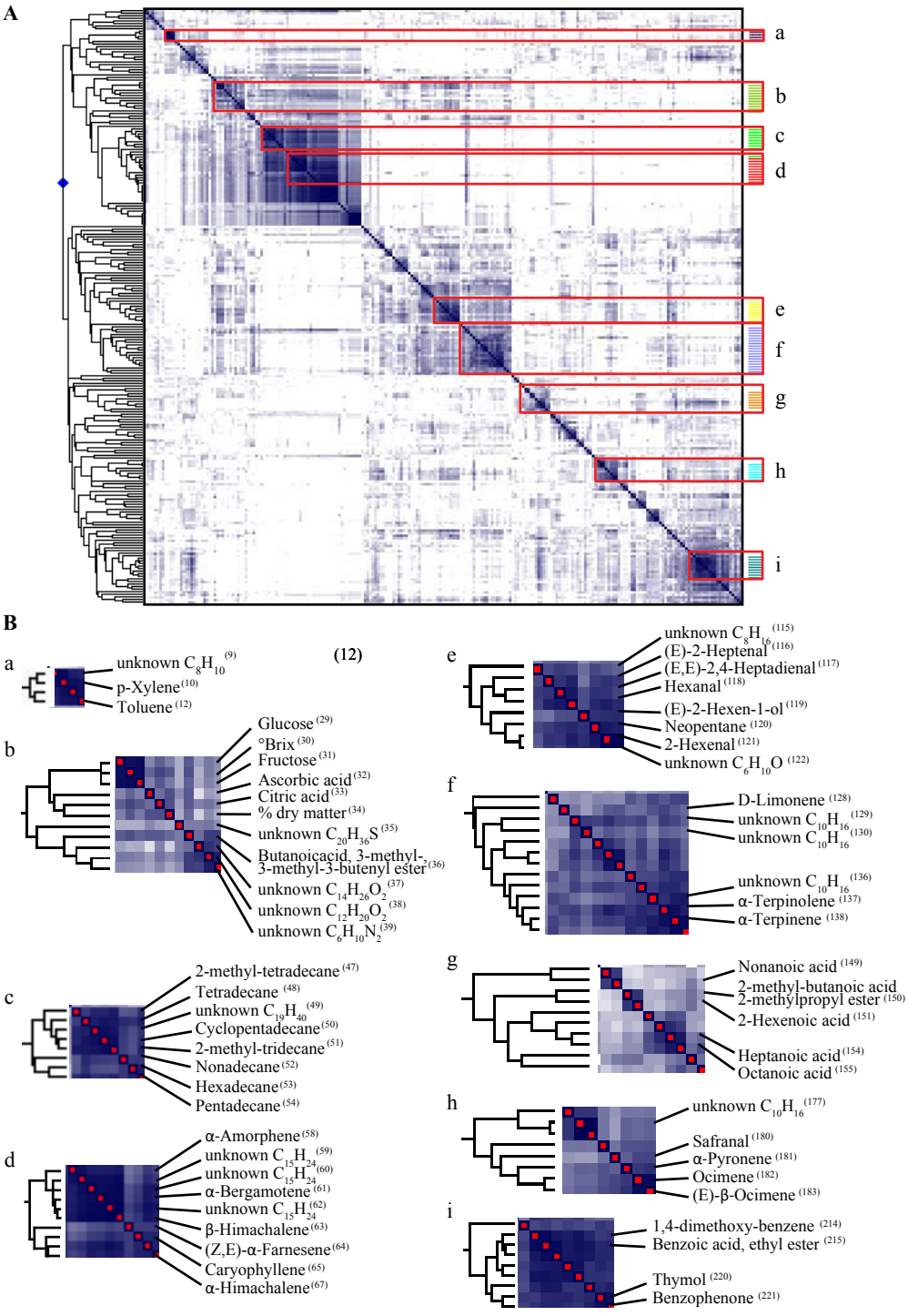

A
a
b
c
d
e
f
g
h
i
B
a
unknown C8H10 (9)
p-Xylene (10)
Toluene (12)
(12)
b
Glucose (29)
°Brix (30)
Fructose (31)
Ascorbic acid (32)
Citric acid (33)
% dry matter (34)
unknown C20H36S (35)
Butanoicacid, 3-methyl-, 3-methyl-3-butenyl ester (36)
unknown C14H26O2 (37)
unknown C12H20O2 (38)
unknown C6H10N2 (39)
c
2-methyl-tetradecane (47)
Tetradecane (48)
unknown C19H40 (49)
Cyclopentadecane (50)
2-methyl-tridecane (51)
Nonadecane (52)
Hexadecane (53)
Pentadecane (54)
d
α-Amorphene (58)
unknown C15H24 (59)
unknown C15H24 (60)
α-Bergamotene (61)
unknown C15H24 (62)
β-Himachalene (63)
(Z,E)-α-Farnesene (64)
Caryophyllene (65)
α-Himachalene (67)
e
unknown C8H16O (115)
(E)-2-Heptenal (116)
(E,E)-2,4-Heptadienal (117)
Hexanal (118)
(E)-2-Hexen-1-ol (119)
Neopentane (120)
2-Hexenal (121)
unknown C6H10O (122)
f
D-Limonene (128)
unknown C10H16 (129)
unknown C10H16 (130)
unknown C10H16 (136)
α-Terpinolene (137)
α-Terpinene (138)
g
Nonanoic acid (149)
2-methyl-butanoic acid 2-methylpropyl ester (150)
2-Hexenoic acid (151)
Heptanoic acid (154)
Octanoic acid (155)
h
unknown C10H16 (177)
Safranal (180)
α-Pyronene (181)
Ocimene (182)
(E)-β-Ocimene (183)
i
1,4-dimethoxy-benzene (214)
Benzoic acid, ethyl ester (215)
Thymol (220)
Benzophenone (221)

Figure 1 (Left). Metabolite-metabolite correlation matrix based on intensity patterns of 224 volatiles, concentrations of 5 primary metabolites plus TSS and dry matter content. **A,** The main compound clusters are situated along the diagonal axis (clusters a-i). Correlations between metabolites are shade colored: the darker the color the higher the percentage of similarity between metabolite expression patterns. **B,** Detailed dendrograms of each compound cluster with putative compound identity. Numbers between brackets refer to the order in Supplementary Table 1. Compound clusters: a and i, phenolics; b, non-volatiles; c, higher alkanes; d, sesquiterpenes; e, lipid derivatives; f and h, terpenoids; g, saturated fatty acids.

Metabolic contrasts

HCA on the three individual repetitions of all genotypes (cv. Vania was only represented twice due to poor quality of the fruits) was performed based on intensity patterns of the 224 volatiles, concentrations of the 5 primary metabolites as well as TSS and dry matter content. The HCA plot (Fig. 2A) showed that repetitions of the same genotype generally clustered together.

Principal components analysis (PCA) revealed two major types of metabolic contrasts within the 24 pepper genotypes (Fig. 2B). First, PCA showed a clear separation between the gene bank accession PBC1405 and all other cultivated and elite genotypes. Secondly, PCA revealed the variation in metabolite content within the cultivated and elite genotypes, separating the yellow blocky genotypes F, G and I from the red conical genotypes N and Piquillo. The third, fourth and fifth principal component (PC) accounted for approximately 5% explained variance each. Adding PCs however did not obviously contribute to further separation (with biological relevance) of the genotypes.

The separate grouping of PBC1405 in the PCA plot was mainly caused by a higher relative abundance of 20 volatiles in PBC1405 compared to the other genotypes. Eight volatiles of those were solely detected in PBC1405: hexanoic acid hexyl ester (83), β-Ionone (55), propanoic acid, 2-methyl-, 1-methylbutyl ester (79), 3,7-dimethyl-6-Octen-1-ol formate (72), (Z)-1,1,3,5-tetramethyl-cyclohexane (75), 3,3-dimethyl-cyclohexanol (70), 2,4-dimethyl-3-pentanone (84) and 1-tridecene (73) (numbers between brackets refer to the order in Supplementary Table 1). Of the other twelve volatiles, trace amounts could be detected in fruits of also a few non-PBC1405 genotypes. Nine volatiles, from the twenty making the difference between PBC1405 and the other genotypes, belonged to the higher alkanes and sesquiterpenes (cluster c/d, Fig. 2C). The metabolites which were most discriminative between yellow genotypes F, G and I versus all other genotypes, were mainly of phenolic origin (cluster i, Fig. 2C): copaene (213), 1,4-dimethoxy-benzene (214), benzoic acid, ethyl ester (215), 2-ethyl-1-hexanol (216), 2-octanone (218), camphor (219), thymol (220) and benzophenone (221). In addition, although not phenolic

derived, 3-hepten-2-one (223) was located within phenolic cluster i (Fig. 2C). The fourth yellow genotype H, remarkably, did not cluster with the yellow genotypes F, G and I, but positioned among red and orange genotypes. Lipid derived volatiles (cluster e, Fig. 2C), including (E)-2-Heptenal (116), (E,E)-2,4-Heptadienal (117), Hexanal (118), (E)-2-Hexen-1-ol (119), Neopentane (120) and 2-hexenal (121), were principally responsible for the separation of the red conical genotypes N and Piquillo.

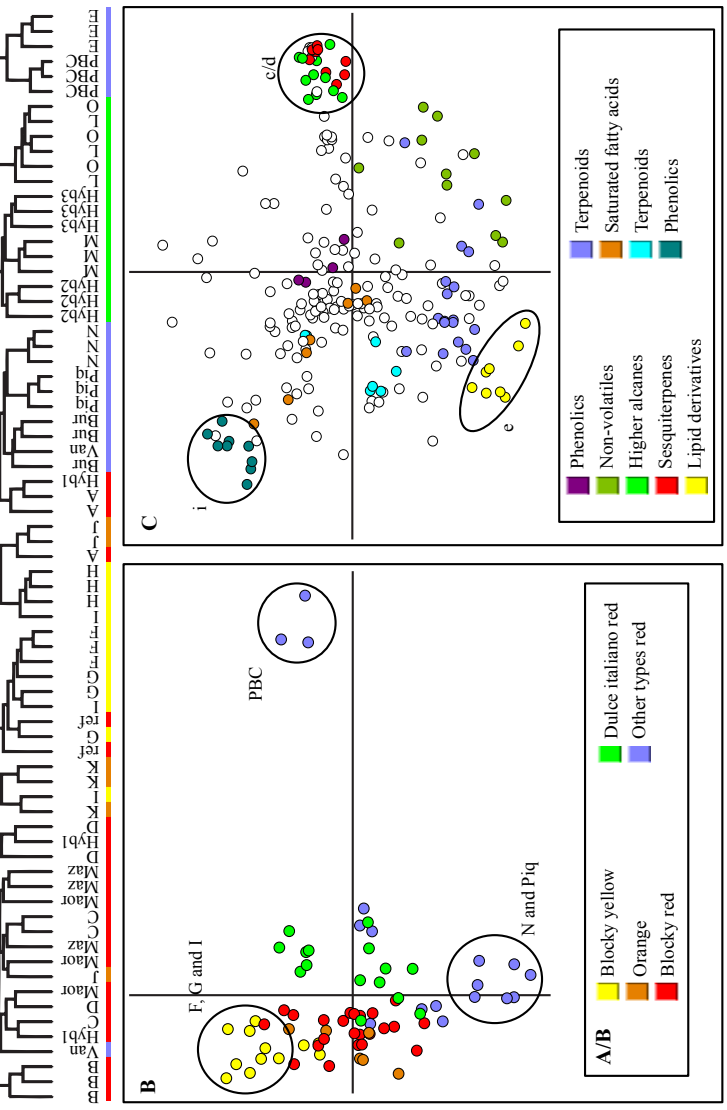


Figure 2. Multivariate analysis of 24 pepper genotypes in 3 replicates and the reference (ref) used in the sensory analysis. **A**, Hierarchical tree of the genotypes based on intensity patterns of 224 volatiles, concentrations of 5 primary metabolites plus TSS and dry matter content. **B**, PCA scores plot showing the major types of differences between all genotypes: discriminating PBC1405 (PBC) from the cultivated/elite genotypes along the horizontal axis (18.3% explained variance) and within cultivated/elite genotype variation along the vertical axis (12.7% explained variance), separating genotypes F, G and I from N and Piquillo (Piq). **C**, PCA loadings plot showing the distribution of the 231 metabolites. Colors of the metabolite clusters correspond with Figure 1.

Sensory evaluation

In addition to the metabolic profiling, the 24 *Capsicum annuum* accessions in this study were evaluated by a trained descriptive sensory panel. The expert panel scored14 attributes on a scale from 0 to 100 to describe the taste sensation in the mouth/ throat. The smell of the fruits was not evaluated separately. Mean attribute values were calculated per replicate of a genotype after correction for session and panelist effects and removal of outliers (Supplementary Table 2). Boxplots of these mean attribute values visualize the variation between attributes (Fig. 3). Clear differences between attributes were observed with respect to median and range. Especially taste attributes green bean, petrochemical, carrot, musty, perfume and to a lower extent also grassiness obtained on average low scores within a limited scoring range. For all other attributes more variability between genotypes was found by the sensory panel which is reflected by higher median values and wider ranges.

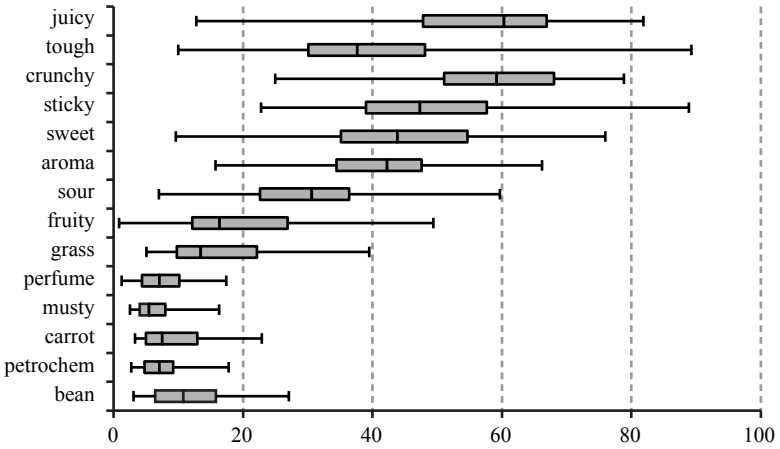


Figure 3. Boxplots of sensory attribute values from 24 pepper genotypes, corrected for session, panellist and outlier effects.

The sensory data was subjected to principal components analysis to describe the relations between the different attributes and to reduce the complexity of the data. The first two principal components explained nearly 70% of the variation in the data, with the first component alone accounting for 52.6% and the second component accounting for 17.2% of the variation (Fig. 4). PCA revealed two sensory contrasts within the pepper genotypes (Fig. 4B), *i.e.* a juiciness versus stickiness/toughness contrast (vector 1) and secondly a sweet/fruity versus sour/grassy contrast (vector 2). The five attributes carrot, bean, petrochemical, musty and perfume taste, which

are situated close to the origin hardly contributed to visualize the genotype variation (Fig. 4A). This is in accordance with their small estimated genotype effects (Table 3). The contrast along vector 1 is texture related, whereas vector 2 mainly describes the basic sweet-sour taste contrast.

In our material vector 1 distinguished the yellow blocky genotype F from the red conical cultivar Piquillo, mainly based on their physical related texture differences, with F having juicy fruits and Piquillo having very tough fruits with a sticky skin (Supplementary Table 2). The dulce italiano genotypes Hybrid 2, 3 and M were characterized by high sweetness scores and were separated along vector 2 from the gene bank accession PBC1405 with sour and grassy fruits (Supplementary Table 2). The positioning of genotypes F, Piquillo and PBC1405 in the extremes of the sensory PCA plot (Fig. 4) was also reflected by their extreme positions in the PCA plot showing the metabolic differences (Fig. 2).

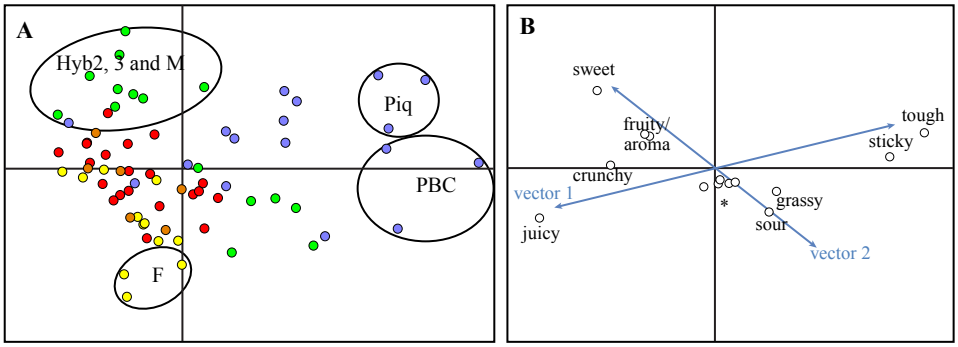


Figure 4. Principal components analysis bi-plot based on sensory data of 24 pepper genotypes in 3 replicates and the reference used in the sensory analysis. **A**, PCA scores plot showing the major types of differences between the pepper genotypes. The extreme dulce italiano genotypes hybrid 2, 3 and M, the yellow blocky genotype F, cv. Piquillo (Piq) and PBC1405 (PBC) are indicated. **B**, PCA loadings plot showing the distribution of the 14 sensory attributes along the horizontal axis (52.6% explained variance) and the vertical axis (17.2% explained variance). Vector 1 indicates the juicy-tough contrast and the sweet-sour contrast is indicated by vector 2. * the five attributes close to the origin are carrot, bean, petrochemical, musty and perfume taste. Colors of the genotypes correspond with the legend of Figure 2.

Variance components of the sensory attributes were estimated using a variance components model with the terms genotype, replicate, their interaction, and sessions within genotype/replicate combinations and taster within sessions (Table 3). In this way it was possible to compare the effects of genotype, repetition, session and taster on the individual attributes. The percentage variance explained by genotype is an estimation for heritability of the attribute. Large differences in genotype effect were observed

between the attributes, ranging from 0 47.1% explained variance for musty, carrot, petrochemical and green bean taste versus juiciness, respectively. Logically, session and panelist effects were largest in the former group of attributes. The attributes grassiness, green bean, carrot, perfume, petrochemical and musty taste, for which the highest number of outlier scores (3-9) were found, were also the attributes with smallest heritability (zero or close to zero; Table 3). On the other hand low heritabilities (10-20%) were found for the attributes sourness, aroma and stickiness, moderate genotype effects (20-35%) for the attributes crunchiness, sweetness and toughness and even a high heritability (47%) for juiciness. In addition the correlation of attributes with total yield was calculated since this is an important breeding parameter (Table 3). In general low correlations were found between taste attributes and yield, whereas moderate correlations were found for the texture related attributes.

Relation between biochemical composition and attributes

A Random Forest (Breiman 2001) regression approach was used to relate each sensory attribute to the volatile (intensity patterns of the 224 volatiles) and non-volatile (concentrations of the 5 primary metabolites as well as TSS and dry matter measurements) data and to determine importance of the individual compounds. The performance of the models is expressed by the prediction R^2, which

Table 3. Estimated variance component percentages of the outlier corrected sensory attributes with genotype (gen), replicate (rep), their interaction, and sessions (sess) within genotype/replicate combinations and taster (tast) within sessions.

Variance comp. %	juicy	tough	crunch	sticky	sweet	aroma	sour	fruity	grass	perf[1]	musty	carrot	petr[2]	bean
Genotype	47.1	32.6	23.1	17.4	25.8	12.2	10.1	9.3	2.7	1.3	0	0	0	0
Repetition	3.8	3.7	1.1	5.2	3.6	1.8	0.3	0.0	0	1.4	1.0	6.3	0	3.8
Gen*rep	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gen*rep*session	0	0	0	0	0	0	4.1	3.6	0	0	0	0	0	0
Gen*rep*taster	14.8	35.8	19.0	56.7	30.0	56.2	34.1	47.2	39.6	35.3	20.1	66.5	39.0	35.8
Gen*rep*sess*tast	33.7	28.0	56.9	20.7	40.6	29.8	51.4	39.9	57.7	62.0	78.9	27.1	61.0	60.4
Corr. total yield	0.51	-0.47	0.25	-0.34	-0.06	-0.05	-0.09	0.02	-0.07	0.08	0.15	0.15	-0.04	-0.07

[1]perfume, [2]petrochemical

to

is not the goodness of fit, like in linear regression, but the percentage of explained variance of samples which were not used to fit the Random Forest model (prediction of the out-of-bag samples). The attributes juiciness, toughness, crunchiness and stickiness as well as the attributes sweetness, aroma, sourness and fruity/apple taste with (reasonably) good heritabilities (Table 3), could be significantly predicted with (combined) volatile and non-volatile data (Table 4). The prediction of these attributes by the individual non-volatile and volatile data was in general as good as the combination of both datasets, with on average the non-volatile data alone (46.3%) performing slightly better than the volatile data alone (42 %; Table 4). The R^2 values of the attributes with smallest heritabilities (zero or close to zero; Table 3), grassiness, perfume, musty, carrot, petrochemical and green bean taste, were not significant ($P>0.05$ in permutation test).

The eight attributes which were significantly predicted (Table 4) could be divided into two groups: (1) texture related attributes: juiciness, toughness, crunchiness and stickiness; and (2) more intrinsic taste related attributes: sweetness, aroma, sourness and fruity/apple taste. As physical properties of the fruits other than dry matter content, like firmness of the pericarp, were not measured, only compounds with significant contribution ($P<0.05$ in permutation test) to the attributes aroma, fruity/apple taste, sweetness and sourness are reported in Table 5. Importance of the individual compounds in the attribute models is expressed by the increase in mean square error (%IncMSE) and where possible the known flavor description of found compounds is given (Table 5).

The lipid derivatives (E)-2-hexen-1-ol and neopentane (cluster e, Fig. 1) as well as the primary metabolites fructose and/or glucose (cluster b, Fig. 1) positively contributed to the prediction of the attributes aroma, fruity/apple taste and sweetness (Table 5). Previously, the flavor of (E)-2-hexen-1-ol has been described as almond, fruit, spicy by sniffing port analysis of fresh bell peppers (Luning et al. 1994a), which is in line with our findings. No flavor description could be found for neopentane. (Z)-linalooloxide, with floral, green bell pepper notes (Luning et al. 1994a) contributed (correlation 0.40) specifically to the prediction of aroma, although to a somewhat lower extent. For fruity/apple taste, the spicy, herbal compound p-menth-1-en-9-al (http://www.thegoodscentscompany.com) was found to be positively correlated, which was not reported in pepper before. Also for sweetness three new pepper flavor related volatiles were found; (E)-β-ocimene, (Z)-2-penten-1-ol and (E)-geranylacetone, and 3 compounds with unknown identity (Table 5). (Z)-β-ocimene was described by the sniffing panel of Luning (Luning et al. 1994a) as rancid, sweaty, however the terpenoid (E)-β-ocimene found in our study (cluster h, Fig. 1) to be positively correlated to sweetness has a sweet, herbal flavor according to the

Good Scents Company. The lipid derivative (Z)-2-penten-1-ol has rubber, plastic, green flavor notes (http://www.flavornet.org/d_odors.html), which does not seem to match with the attribute sweetness, however this might be explained by their negative correlation (Table 5), meaning that a higher (Z)-2-penten-1-ol concentration could mask sweetness (and vice versa). Finally, (E)-geranylacetone was found to be correlated in a positive manner with sweetness, which is in line with its floral, fruity, pear, apple and banana with tropical nuances flavor description (Mosciano 1998). For sourness only an unknown $C_6H_8O_2$ compound was found which significantly contributed to its prediction, with a direct correlation of 0.38. Remarkably, the same compound was found to be predictive for sweetness, however in that case with a negative correlation.

Table 4. Random Forest prediction R^2 values of the attributes which could be significantly predicted by the volatile (vol) data, the non-volatile (nonvol) data and their combination (volnonvol).

Attribute	Volnonvol	Nonvol	Vol
Juicy	71.1	63.5	67.4
Tough	65.1	65.3	61.9
Crunchy	34.1	47.0	33.5
Sticky	54.3	41.5	53.2
Sweet	54.6	56.1	47.8
Aroma	42.7	39.0	41.3
Sour	14.5	21.7	10.3
Fruity	31.8	35.9	20.5
Average	46.0	46.3	42.0

Discussion

Metabolic contrasts between genotypes are caused by clusters of volatile and non-volatile compounds

We investigated the taste of 24 non-pungent *Capsicum annuum* genotypes in relation to their biochemical profile in this study. In accordance to the genetic diversity, the genotypes displayed a high degree of fruit type, organoleptic and metabolic variation. A total of 224 different volatile compounds could be distinguished using GC-MS in combination with multivariate mass spectral reconstruction. This number of volatiles was very comparable to the number of volatiles found in twelve *C. annuum* genotypes by Rodriguez-Burruezo and colleagues (2010). Since non-volatiles are

also important determinants of sensory perceived taste, the concentration of sugars and acids was determined in addition to the GC-MS measurements. The sucrose concentration in ripe fruits turned out to be under the detection limit (0.3 g/100 g fresh weight) of our enzymatic determination method, although it was detected at low concentration in green, turning and red fruits of Mazurka (0.31, 0.65 and 0.19 g/100g fresh weight) by Luning (1994b). For both volatile and non-volatile compounds, highly correlated clusters were found by HCA, which could be related to metabolic pathways and common biochemical precursors. The specific grouping of the non-volatiles glucose, fructose, citrate and ascorbic acid with the volatile compound 3-methyl-butanoic acid 3-methyl-3-butenylester and several other volatiles of unknown identity (cluster b, Fig. 1) seemed rather caused by population structure than by functional relationship. Metabolic contrasts between genotypes were caused by both qualitative and quantitative differences in the metabolic clusters, with the phenolic derivatives, higher alkanes, sesquiterpenes and lipid derived volatiles forming the major determinants. Changes of genes (expression) in such pathways would probably change complete clusters of volatiles, thereby affecting individual attributes or even overall flavor (e.g. Lewinsohn et al. 2005, Tieman et al. 2006).

3-hepten-2-one: a candidate carotenoid degradation product?

Separate clustering of the three yellow genotypes F, G and I from the other orange and red genotypes by PCA (Fig. 2C) suggested a relation with fruit color. However, the fourth yellow genotype H positioned among red and orange genotypes and the metabolites, which were most discriminative between both groups, were mainly of phenolic origin and could therefore not be linked to fruit color. We do nevertheless know, that pepper fruit color is caused by specific carotenoids from which several volatiles, like 6-methyl-5-hepten-2-one or β-ionone, are derived (Krammer et al. 2002). In our material, however, this class of carotenoid derived volatiles was only found to a very limited extent, with e.g. 6-methyl-5-hepten-2-one only formed at low intensity in the red blocky genotype B (188, Supplementary Table 1). By contrast, the $C_7H_{12}O$ compound 3-hepten-2-one (223), with a very similar chemical structure as 6-methyl-5-hepten-2-one, was widely present in our panel. PCA revealed that 3-hepten-2-one, although not of phenolic origin, located within the phenolic cluster i (Fig. 2C), which separated the yellow genotypes from the orange and red ones. Comparing the intensity of 3-hepten-2-one between genotypes with different fruit colors, 3-hepten-2-one was indeed found to be on average 1.6 times higher in the orange genotypes and even 1.9 times higher in the yellow genotypes versus the red fruited genotypes. These higher expression levels are comparable with measurements in an

orange tomato mutant *tg*, which accumulates the orange pigment prolycopene and in which 3-fold higher levels of 6-methyl-5-hepten-2-one were found in comparison to the wild-type red tomato (Lewinsohn et al. 2005). Based on its chemical structure and expression pattern it seems therefore interesting to study in more detail whether 3-hepten-2-one, like 6-methyl-5-hepten-2-one, can be a true carotenoid degradation product, explaining its higher intensity in yellow and orange fruits.

Table 5. Compounds with significant contribution in the attribute prediction.

Attribute	Compound[1]	Cl[2]	%IncMSE[3]	Corr[4]	Flavor description	Reference
Aroma	(E)-2-Hexen-1-ol (119)	e	31.1	0.60	Almond, fruit, spicy	Luning et al. 1994a
	fructose (31)	b	15.9	0.60	Sweet	
	Neopentane (120)	e	14.2	0.45	-	
	Brix (30)	b	3.4	0.49	-	
	(Z)-Linalooloxide (100)	-	2.9	0.40	Floral, green bell pepper	Luning et al. 1994a
Fruity	(E)-2-Hexen-1-ol (119)	e	5.1	0.52	Almond, fruit, spicy	Luning et al. 1994a
	fructose (31)	b	4.6	0.56	Sweet	
	Neopentane (120)	e	3.9	0.35	-	
	glucose (29)	b	2.9	0.52	Sweet	
	p-Menth-1-en-9-al (102)	-	2.2	0.52	Spicy, herbal	Good scents co.[5]
Sweet	(E)-2-Hexen-1-ol (119)	e	71.4	0.54	Almond, fruit, spicy	Luning et al. 1994a
	glucose (29)	b	26.6	0.61	Sweet	
	Neopentane (120)	e	25.9	0.41	-	
	fructose (31)	b	16.5	0.65	Sweet	
	(E)-β-Ocimene (183)	h	6.9	0.22	Sweet, herbal	Good scents co.
	unknown $C_{15}H_{24}$ (28)	-	3.6	-0.25	-	
	unknown $C_{10}H_{16}$ (181)	h	3.6	0.34	-	
	unknown $C_6H_8O_2$ (190)	-	3.3	-0.29	-	
	(Z)-2-penten-1-ol (27)	-	3.2	-0.10	Rubber, plastic, green	Flavornet [6]
	(E)-Geranylacetone (203)	-	2.7	0.35	Floral, fruity, pear, apple and banana with tropical nuances	Mosciano 1998
Sour	unknown $C_6H_8O_2$ (190)	-	2.7	0.38	-	

[1] Numbers between brackets refer to the order in Supplementary Table 1. [2] Compound cluster (Fig. 1). [3] Percentage increase in mean square error after permutation. [4] Pearson's correlation coefficient based on log transformed metabolic data. [5] http://www.thegoodscentscompany.com , [6] http://www.flavornet.org/d_odors.html.

A large role for texture and sweetness/sourness in pepper flavor

The variation in taste could be reduced into two major contrasts, which were a texture related contrast and the basic sweet-sour contrast. In tomato a similar sweet/fruity versus sour/watery contrast has been found in both a study with 16 tomato cultivars (Sinesio et al. 2010) and a study with 94 tomato varieties (Hageman et al. 2010), with a texture related contrast in the second principal component, which for tomato described a firmness-mealiness contrast. Although we found similar sensory contrasts as in tomato, in pepper the texture contrast explained the largest part of variation (52.6%), whereas in tomato the sweet-sour contrast was most discriminative (~45% explained variation).

Towards a general pepper taste model

Physical properties of the fruits, like firmness and flexibility of the fruit flesh or skin, were generally not measured. So far the only measured physical trait was dry matter content, which correlated well with the texture related attributes stickiness (0.57), toughness (0.65) and juiciness (-0.61). The correlation with crunchiness was much smaller (-0.15). To complement our study it would therefore be interesting to perform physical fruit measurements in relation to the texture attributes. Additionally, consumer liking data of our genotypes would make it possible to predict overall flavor (liking) instead of individual attribute prediction. Currently we are working on the development of such a general pepper taste model within a Dutch research consortium (TTI-GG).

Aroma, fruity/apple taste and sweetness can be well predicted by metabolites

Expected relations between (non-)volatiles and attributes, like sweetness and sugars, were found but also some new relations. Neopentane was found to contribute to both aroma and fruity/apple taste as well as sweetness, however no flavor description could be found. The most likely explanation for neopentane being correlated was that it has a very similar expression as (E)-2-hexen-1-ol (Fig. 1, cluster d), which seemed truly predictive based on its almond, fruit, spicy odor description (Luning et al. 1994a). For fruity/apple taste and sweetness new relations were found with the compounds p-menth-1-en-9-al, (E)-β-ocimene, (Z)-2-penten-1-ol and (E)-geranylacetone. Taking both the flavor description of these compounds and the direction of their correlation with either fruity/apple taste or sweetness into account, it seemed reasonable that all four compounds are really contributing to the involved

attribute. In our analyses we did not find an effect on flavor attributes of the well known compound 2-isobuthyl-3-methoxypyrazine, which is commonly described in sniffing port analyses as characteristic (green) bell pepper aroma (Luning et al. 1994a, Van Ruth et al. 1995, Rodriguez-Burruezo et al. 2010), indicating different sensitivity of sniffing versus taste evaluations. It should also be emphasized that the lists of compounds with significant contribution in attribute prediction (Table 5) do not contain all compounds with high correlation to the attributes. This is due to the Random Forest multiple linear regression technique we used. For this reason there are more (often correlated) compounds contributing to the attributes, but only the best one or two predictors from such compound clusters are listed. For example, in the case of aroma, in addition to (E)-2-hexen-1-ol and neopentane, also 2-hexenal from the same lipid derivative cluster (cluster e, Fig. 1) is strongly correlated (0.52) to aroma, but not listed in table 5. Overall it can be concluded that sugars and several lipid derivates from cluster e (Fig. 1) as well as (Z)-2-penten-1-ol play a big role in pepper flavor determination, influencing at least the attributes, aroma, fruity/apple taste and sweetness.

Perspectives for flavor improvement by breeding

As mentioned, several attributes could be predicted by volatile and/or non-volatile compounds, making the discussed compounds interesting targets for breeding. Attributes like, grassiness, perfume, musty, carrot, petrochemical and green bean taste could however not significantly be predicted, which was caused by small differences between genotypes and sometimes large variation within genotypes, resulting in low heritabilities. Variation in these attributes seems to be caused more by environmental variation and variation in panellists' evaluations rather than by strong genetic effects. From the eight significantly predicted attributes, sourness had the lowest Random Forest prediction R^2 value of only 14.5%, just above the significance threshold found by permutation. The only compound with a significant contribution to sourness was an unknown $C_6H_8O_2$ compound. From the organic acids, citrate showed the best relation with sourness with a correlation of 0.34, which was however not significant. Organic acids seem therefore not to play a role of importance in pepper sourness. In tomato, however, a correlation of 0.76 was found between titratable acids (mainly citric and malic acid) and sourness (Tandon et al. 2003), while the variation and concentration of organic acids in that study were similar as in our pepper collection. An explanation for the low prediction R^2 of sourness could be that the effect of sour related metabolites is masked by other volatile and non-volatile compounds or texture differences. This complex nature and its low heritability (10.1%) make

sourness a difficult attribute to influence by breeding. Attributes, like sweetness and the texture attributes, showing high heritability and a wide scoring range are on the other hand very amenable traits for improvement by breeding. Moreover, based on the genotypes in our panel it seems possible to increase juiciness, crunchiness and sweetness in pepper fruits without (too much) negative influence on total yield. Another way to enrich commercial breeding programs for flavor variation would be the use of gene bank material, since in our case, gene bank accession PBC1405 was extreme in both the metabolic and sensory analyses. With this study we made significant progress in the understanding of pepper taste. This may eventually lead to more effective breeding strategies towards pepper varieties with improved taste.

Acknowledgements

The authors kindly acknowledge Harry Jonker and Yvonne Birnbaum at Plant Research International for performing the SPME-GC-MS analyses. In addition we thank Suzanne de Wit and Gerald Freymark for technical support and useful discussions. Finally we are grateful to Laure Flament, Femke Willeboordse, Sander Bos, Tineke Benning and Paula de Grauw for performing a massive job on sample preparation and non-volatile measurements, and all other people at Rijk Zwaan who took care of perfect greenhouse management.

Supporting information available

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.foodchem.2011.10.081.

Chapter 4

Prediction of sweet pepper (*Capsicum annuum*) flavor over different harvests

P.M. Eggink, C. Maliepaard, Y. Tikunov, J.P.W. Haanstra,
L.M.M. Pohu-Flament, S.C. de Wit - Maljaars, F. Willeboordse - Vos,
S. Bos, C. Benning - de Waard, P.J. de Grauw - van Leeuwen,
G. Freymark, A.G. Bovy, R.G.F. Visser
Euphytica (2012) 187: 117-131

Abstract

To better understand and predict the complex multifactorial trait flavor, volatile and non-volatile components were measured in fresh sweet pepper (*Capsicum annuum*) fruits throughout the growing season in a diverse panel of twenty-four breeding lines, hybrids, several cultivated genotypes and one gene bank accession. Biochemical profiles were linked to individual flavor attributes, that were objectively quantified by a trained descriptive expert panel. We used a Random Forest regression approach for prediction of the flavor attributes within and between harvests. Predictions of texture related attributes (juiciness, toughness, crunchiness and stickiness of the skin) and sweetness were good (around 60-65% in the analyses with the three harvests combined). The predictions of the attributes aroma intensity, sourness and fruity/apple were somewhat lower and more variable between harvests. (E)-2-hexen-1-ol, neopentane, p-menth-1-en-9-al, 3-hepten-2-one, (Z)-β-ocimene, (Z)-2-penten-1-ol, 1-methyl-1,4-cyclohexadiene, glucose, fructose and three unknown volatile compounds were identified as key-metabolites involved in the flavor differences between both genotypes and harvests. The complex nature of flavor is exemplified by the observed masking effect of fructose and other sugars on sourness and sourness related metabolites, like citrate. The knowledge obtained from the overall biochemical, sensory and prediction analyses forms a basis for targeted flavor improvement by breeding.

Keywords: biochemical profiling, flavor, multivariate prediction, Random Forest, sensory evaluation.

Introduction

Pepper fruits (*Capsicum annuum*) are widely consumed either as a fresh vegetable or dried as a spice. Especially, sweet peppers are eaten fresh or processed, as unripe (green or white) or ripe (e.g. red, yellow and orange) fruits. Fruits are commonly used in diets because of their typical color, pungency, taste and/or distinct aroma (Govindarajan 1985). In the breeding process of pepper, flavor has been identified as a chance for added value creation and consumer satisfaction (Verkerke 2000). However, flavor is a complex multifactorial trait (combination of taste, aroma, mouth feel, sight and sound) and in case of a biological product like pepper, is influenced by environmental factors during growth and post-harvest processing. As a consequence, until now it has been difficult to measure pepper flavor in a high-throughput and quantitative way. So far, only some studies have been performed to understand the

role of various volatile and non-volatile components in fresh pepper flavor (e.g. Luning et al. 1994a, Rodriguez-Burruezo et al. 2010), but a clear understanding of the interaction between them and their effect on flavor is lacking.

In the present study we characterized flavor over different harvests in a broad germplasm panel from a commercial breeding program completed with a few cultivated genotypes and a gene bank accession. Flavor was objectively quantified by thorough biochemical profiling in combination with sensory evaluations by a trained expert panel. The overall results make it possible to link individual flavor attributes to volatile and non-volatile components and to develop models for prediction of sensory descriptors over harvests. Our results form a starting point for a better understanding of pepper flavor and targeted improvement by breeding.

Materials and methods

Plant material

In this study the 24 non-pungent *Capsicum annuum* accessions described by Eggink and colleagues (2012a) were used (Table 1). In short, the genotype panel consisted of elite pepper breeding lines and hybrids (provided by the breeding company Rijk Zwaan), several cultivated genotypes (landraces and old hybrids) and one gene bank accession. In 2008, the genotypes were grown in soil in a greenhouse at Rijk Zwaan (De Lier, The Netherlands), according to standard Dutch pepper management conditions with 2.5 plants/m^2. Potential shading effects, because of the diverse nature of the genotypes, were avoided by ordering the plants by (expected) plant height in the greenhouse in 3 separate blocks (i.e. tall, intermediate and short plants). All genotypes were grown in 3 plots of 5 plants, which were randomized within the separate blocks.

From the beginning of May till the end of September, all completely (95-100%) colored fruits were harvested, counted and weighed on a (bi)weekly base. In that period 3 harvests (29 May, 31 July and 4 September) were used for biochemical measurements and sensory evaluation.

After harvesting, fruits were stored in a climate room at 20°C with 80% relative humidity for 4-5 days to optimize ripening. This is standard procedure to mimic the Dutch commercial system. From each individual repetition of the genotypes, a selection of 5-8 fruits was pooled to make a representative fruit sample. Fruits were washed with cold tap running water, dried with a clean towel, cut (top and bottom parts were discarded) in 1-2 cm pieces, mixed and seeds were removed. Half of the fruit pieces from each sample were immediately frozen in liquid nitrogen,

ground in an electric mill and stored at -80°C while the other half was used for flavor evaluation.

Table 1 Description of *Capsicum annuum* genotypes evaluated for fruit quality attributes

Genotype	Origin	Source country	Fruit type	Size[a] (cm)	Color	Yield[b] (kg/m^2)
Mazurka[c]	Elite	Netherlands	Blocky	8 x 8	Red	12.1 ± 0.1
Hybrid 1	Elite	Netherlands	Blocky	8 x 8	Red	12.8 ± 0.8
Line A	Elite	Netherlands	Blocky	8 x 8	Red	11.7 ± 0.3
Line B	Elite	Netherlands	Blocky	8.5 x 8	Red	9.6 ± 0.9
Line C	Elite	Netherlands	Blocky	8.5 x 8	Red	12.9 ± 0.3
Line D	Elite	Netherlands	Blocky	8 x 8	Red	9.1 ± 0.7
Line F	Elite	Netherlands	Blocky	9 x 8	Yellow	11.8 ± 1.3
Line G	Elite	Netherlands	Blocky	8 x 8	Yellow	15.3 ± 0.6
Line H	Elite	Netherlands	Blocky	8 x 8	Yellow	12.9 ± 0.5
Line I	Elite	Netherlands	Blocky	8 x 8.5	Yellow	14.6 ± 1.4
Line J	Elite	Netherlands	Blocky	8 x 8.5	Orange	12.4 ± 2.4
Line K	Elite	Netherlands	Mini block	5 x 5	Orange	7.2 ± 1.1
Hybrid 2	Elite	Italy	Dulce italiano	20 x 4	Red	10.9 ± 2.0
Hybrid 3	Elite	Italy	Dulce italiano	22 x 4.5	Red	13.0 ± 0.8
Line L	Elite	Italy	Dulce italiano	22 x 4	Red	11.5 ± 0.5
Line M	Elite	Italy	Dulce italiano	18 x 4.5	Red	9.8 ± 1.4
Line O	Elite	Italy	Dulce italiano	22 x 4	Red	11.6 ± 3.0
Line E	Elite	Turkey	Dolma	7 x 6.5	Red	8.7 ± 1.1
Line N	Elite	Turkey	Kapya	12 x 4	Red	8.8 ± 1.1
Piquillo	Cultivated	Spain	Conical	9 x 4	Red	6.6 ± 0.7
Buran	Cultivated	Poland	Lamuyo	10 x 7	Red	11.0 ± 0.3
PBC1405	Gene bank	AVRDC, Taiwan	Elongated	18 x 2	Red	8.8 ± 1.0
Vania	Cultivated	France	Lamuyo	14 x 8	Red	10.1 ± 0.9
Maor	Cultivated	Spain	Blocky	8 x 8	Red	12.0 ± 0.8

[a] Size is indicated by length x width, [b] Average yield and standard deviation in the harvesting period May through September, [c] Standard variety (reference to e.g. Luning et al. 1994a).

Sensory analysis

The descriptive analysis took place in a sensory laboratory at Wageningen UR Greenhouse Horticulture (WUR-GH, Bleiswijk, The Netherlands). Sixteen panelists, who are part of a trained panel with broad experience in sensory evaluations of food products, including pepper, took part in the experiment. In the weeks prior to the test sessions, panelists participated in specific training sessions on the products.

During the training sessions, panelists reached the consensus on fourteen attributes to describe the flavor sensation in the mouth/throat, which were the texture attributes crunchiness, stickiness of the skin, toughness and juiciness, the basic taste attributes sweetness and sourness and the retronasal flavor attributes aroma intensity, grassiness, green bean flavor, carrot flavor, fruity/apple flavor, perfume flavor, petrochemical flavor and musty flavor. The odor (nasal) of the fruits was not evaluated separately.

During the test sessions, each panelist evaluated all 24 genotypes in a randomized setup, split over 2 subsequent days. On both days, 2 sessions with 6 genotypes and a reference (commercial red blocky *C. annuum* hybrid) were evaluated per panelist. The setup was chosen in such a way that each panelist evaluated one repetition of all genotypes and that each sample was evaluated by 5-6 panelists.

Each panelist received 5 fruit pieces per sample in a ceramic cup and they were asked to mark the intensity of the attributes on a horizontal 100-mm structured line scale on paper. The pepper fruit pieces were swallowed by the panelists. Between samples, panelists rinsed their mouth with tasteless mineral water to neutralize their palate and were also allowed to eat a small unsalted cracker before rinsing their mouth. No information was provided to the panel about the pepper cultivars.

Metabolic profiling

Biochemical profiling was performed as described in Eggink and others (2012a). In short, the fruit samples of the 24 *C. annuum* accessions from the three harvests were analyzed in a single headspace SPME-GC-MS experiment. Derived GC-MS profiles were simultaneously processed by the MetAlignTM software package (http://www.metalign.nl) for baseline correction, noise estimation and ion-wise mass spectral alignment. The Multivariate Mass Spectral Reconstruction (MMSR) approach (Tikunov et al. 2005) was used to reduce data to volatile compound mass spectra. Each compound was represented by a single selective ion fragment in the following multivariate data analysis. The compounds (number of fragment ions in a mass spectrum ≥5) were then subjected to a tentative identification using the NIST mass spectral library (http://www.nist.gov). In total 254 putative volatile compounds were detected of which reliable identities (mass spectra match factor ≥ 700 and identity probability rank ≥ 2, Mihaleva et al. 2009) could be assigned to 129 of them. Identities of 21 volatiles were confirmed by authentic chemical standards (Sigma). Intensities of the 254 volatile compounds which were below the detection limit, in certain genotypes, obtained a random value between 450 and 500. In addition to this, the concentration of sugars (fructose and glucose) was measured by enzymatic determination (Velterop and Vos 2001). Anion exchange

chromatography was used for citric, malic and ascorbic acid determination based on standard protocols (Dionex Corporation, Sunnyvale, CA; http://www.dionex.com/ Application Note 143 "Determination of Organic Acids in Fruit Juices"). Sugar and acid measurements were completed by dry matter content and total soluble solids (brix) determination. Relative intensity patterns of the 254 volatile compounds and the non-volatile chemical composition is given in Online Resource 1.

Data analysis

The sensory data was analysed per harvest in Genstat version 12 using a linear mixed model REML (residual maximum likelihood) analysis with genotype, replicate and their interaction as fixed terms. Sessions (tasting sessions) within replicate/genotype combinations and panelists within sessions were taken as random terms. Mean values were calculated per genotype per replicate after a correction for session and panelist effects and removal of strong outliers (if the absolute value of a standardized residuals was larger than three residual standard deviations) (Online Resource 2).

Principal components analysis (PCA) as implemented in GeneMaths XT version 2.0 was used for visualizing relationships between and among metabolites and attributes. For these analyses the metabolite data sets were log transformed and mean centered. Pearson's correlation coefficient was used as a measure for metabolite-metabolite correlations.

Variance components of the sensory attributes were estimated using a variance components model with REML in Genstat version 12 with the terms genotype, replicate, their interaction, and sessions within genotype/replicate combinations and taster within sessions.

A Random Forest (Breiman 2001) regression approach was used to relate each sensory attribute (response) to the volatile and non-volatile data (predictors) and to determine importance of the individual volatiles and non-volatiles. A double ten-fold cross-validation approach was used to optimize the number of variables for each decision rule in the random forest (the 'mtry' parameter in the R function to perform Random Forest) and to estimate the mean square error (on independent test samples). The performance of the models is expressed by the prediction R^2, which is calculated from the out-of-bag samples (Breiman 2001). This R^2 value therefore is not a goodness-of-fit of the data at hand but an estimate of predictive accuracy on independent (left-out) samples. Variable importance was estimated by the increase in mean square error (MSE) after permutation (Breiman 2001). Significance of the prediction R^2 and variable importance was determined by another permutation test, in which, in turn, each of the sensory attributes was permuted over the genotypes while

retaining the original metabolic values of the genotypes. In this permuted situation, 100 new Random Forest models were constructed with calculation of the prediction R^2 and increase in MSE to estimate variable importance. Significance thresholds were determined at the 95-percentile of the permutations.

Results

Metabolic profiling and sensory analysis

Both the metabolic and the sensory data (Online Resources 1 and 2) from the combined three harvests were subjected to principal components analysis (PCA, Fig. 1). The same extreme genotypes (Fig. 1A,C), metabolite clusters (Fig. 1B) and attributes (Fig. 1D) as described by Eggink et al. (2012a, solely harvest 1 data) are indicated. Repetitions of the same genotype generally group together over harvests, indicated by the circles in Figure 1A and 1C.

An analysis of variance on the eight most contrasting attributes from Figure 1D (juiciness, toughness, crunchiness, stickiness of the skin, sweetness, aroma intensity, sourness and fruity/apple flavor) was performed using a variance components model with the terms genotype, replicate, their interaction, and sessions within genotype/replicate combinations and taster within sessions (Table 2). The percentage variance explained by genotype is an estimation for (broad-sense) heritability of the attribute. Genotype and panelist (by session) effects were largest.

To compare the individual harvests more thoroughly, we first checked for differences in overall attribute scores between harvests by constructing box plot graphs with mean values of the eight most contrasting attributes. The box plots showed harvest differences for all attributes (Fig. 2), which were found to be significant based on a paired T-test ($p<0.05$) . It can be concluded that harvest 3 on average was evaluated as more sweet, less sour and less sticky/tough than harvests 1 and 2. On the other hand, the second harvest could be differentiated from the other two harvests by lower fruity/apple scores and slightly higher crunchiness scores. The first harvest, did not significantly deviate from harvest two and three together for any of the attributes.

Subsequently we checked for harvest differences in the overall metabolic profiles of the 24 pepper accessions, as a starting point for finding metabolites responsible for the differences in flavor (attributes) between harvests. For this purpose the genotypes in the previous principal components analysis (Fig. 1) were colored according to harvesting moment (Fig. 3, green=harvest 1, red=harvest 2 and blue=harvest3). The first two principal components, however, did not show a harvesting moment

effect, as the three harvests completely intertwined (Fig. 3A). Plotting PC7 against PC3 showed, however, that samples from the same harvest, other than genotype repetitions between harvests, grouped together (Fig. 3B). The colored ellipses indicate the separation of harvest 1 samples from the third harvest, while samples from the second harvest are positioned intermediate.

Metabolites which were most contrasting between the harvests are indicated in the green and blue ellipses in Figure 3C. Both groups contained ten metabolites, of which the average mass peak intensity patterns are listed in Table 3. Metabolite contrasts mainly occurred due to quantitative differences between harvests, but in the case of metabolite 740 (unknown $C_5H_{10}O_2$ compound) was caused by a qualitative difference, since the volatile was mainly below the detection limit (<500) in the first harvest. Metabolites in the green ellipse (Fig. 3C) were generally higher in the first harvest than in the third harvest, whereas the opposite was true for the metabolites from the blue ellipse.

Table 2 Estimated variance component percentages of the outlier corrected sensory attributes per harvest with genotype (gen), replicate (rep), their interaction, and sessions (sess) within genotype/replicate combinations and taster (tast) within sessions

Variance comp. %[a]	Juicy	Tough	Crunchy	Sticky	Sweet	Sour	Aroma	Fruity
Genotype (1)	47.1	32.6	23.1	17.4	25.8	10.1	12.2	9.3
Genotype (2)	45.1	38.4	22.8	19.5	28.0	5.3	3.9	2.0
Genotype (3)	41.0	41.4	26.1	22.4	27.3	6.0	13.3	12.1
Repetition (1)	3.8	3.7	1.1	5.2	3.6	0.3	1.8	0.0
Repetition (2)	5.5	1.2	0.0	0.4	1.0	3.4	2.1	2.1
Repetition (3)	1.2	3.5	2.0	0.8	0.7	1.6	2.9	16.1
Gen*rep (1)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gen*rep (2)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gen*rep (3)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gen*rep*session (1)	0.5	0.0	0.0	0.0	0.0	4.1	0.0	3.6
Gen*rep*session (2)	0.6	0.0	3.6	0.0	0.0	1.4	0.0	0.0
Gen*rep*session (3)	0.0	5.0	0.0	1.5	0.4	0.0	0.0	0.0
Gen*rep*taster (1)	14.8	35.8	19.0	56.7	30.0	34.1	56.2	47.2
Gen*rep*taster (2)	26.8	24.6	50.2	52.3	16.7	50.8	33.6	38.1
Gen*rep*taster (3)	24.0	0.0	49.7	21.3	20.8	40.7	16.9	29.6
Gen*rep*sess*tast (1)	33.7	28.0	56.9	20.7	40.6	51.4	29.8	39.9
Gen*rep*sess*tast (2)	21.9	35.8	23.4	27.8	54.3	39.1	60.3	57.8
Gen*rep*sess*tast (3)	33.8	50.1	22.2	54.0	50.8	51.7	66.9	42.2

[a] Harvest number is indicated between brackets.

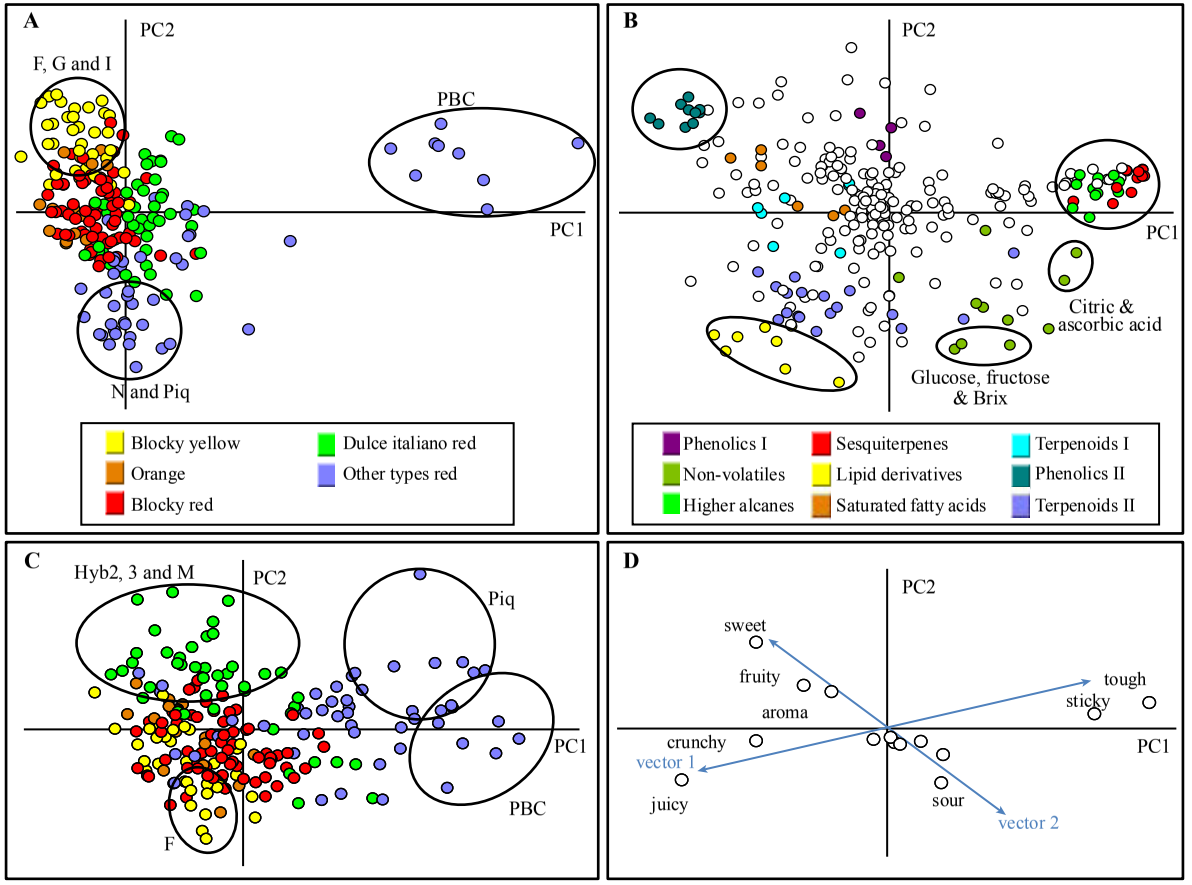


Figure 1. Multivariate analysis on metabolic (**A,B**) and sensory (**C,D**) data of 24 pepper genotypes in 3 replicates of each 3 harvests. **A and C**, PCA scores plots showing the major variation between the genotypes based on (**B**) intensity patterns of 254 volatiles, concentrations of 5 primary metabolites, brix, and dry matter content (PC1=16.2%), PC2=9.5%) and on (**D**) loadings of the 14 sensory attributes (PC1=50.4%), PC2=16.1%)

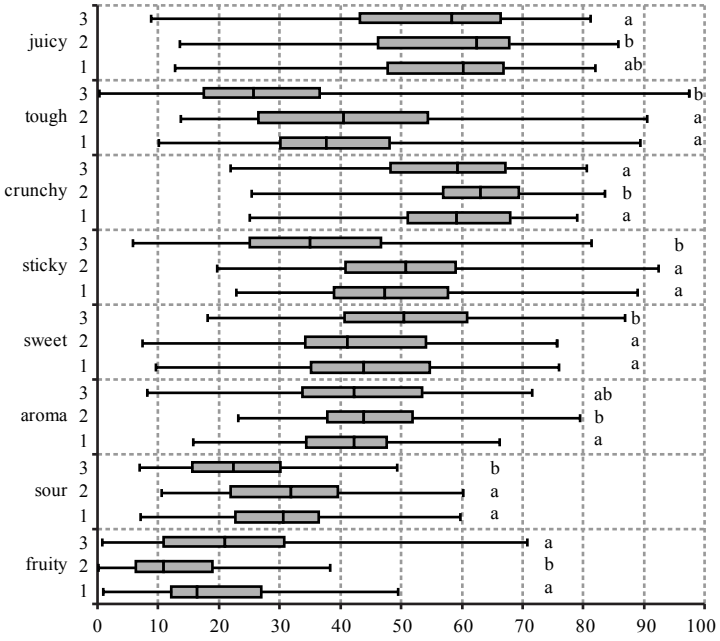


Figure 2. Boxplots of sensory attribute values of 24 pepper genotypes, corrected for session, panelist and outlier effects. The 3 harvests are indicated separately and attributes are divided by horizontal lines. Similar letters between harvests indicate that the overall attribute score is not significantly different (paired T-test, $p<0.05$)

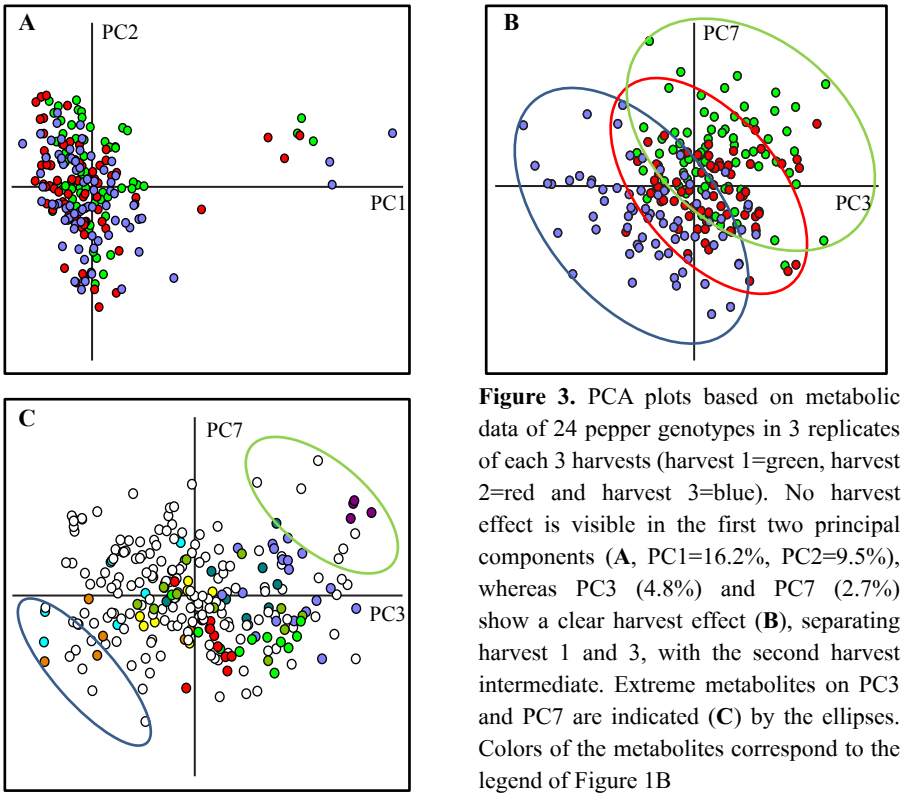


Figure 3. PCA plots based on metabolic data of 24 pepper genotypes in 3 replicates of each 3 harvests (harvest 1=green, harvest 2=red and harvest 3=blue). No harvest effect is visible in the first two principal components (**A**, PC1=16.2%, PC2=9.5%), whereas PC3 (4.8%) and PC7 (2.7%) show a clear harvest effect (**B**), separating harvest 1 and 3, with the second harvest intermediate. Extreme metabolites on PC3 and PC7 are indicated (**C**) by the ellipses. Colors of the metabolites correspond to the legend of Figure 1B

Table 3 Average mass peak intensity patterns of 20 volatile compounds contrasting between the harvests

Compound	Harvest 1	Harvest 2	Harvest 3
Toluene (130)	53074[a]	38279[a]	11317[b]
2-Hexanone (151)	279707[a]	270089[a]	162250[b]
Acetic acid butyl ester (211)	41180[a]	19625[b]	9160[c]
p-Xylene (461)	47312[a]	20096[b]	24774[b]
unknown $C_7H_{14}O$ (471)	32852[a]	32295[a]	21650[b]
Styrene (508)	12748[a]	6542[b]	4589[c]
unknown C_8H_{10} (517)	16312[a]	5447[b]	8595[b]
Hexanoic acid (684)	250125[a]	168540[b]	166638[b]
2,5-dimethyl-2,5-Hexanediol (1438)	52050[a]	24367[b]	1010[c]
unknown $C_9H_8O_2$ (1673)	19625[a]	19752[a]	16059[b]
2-Heptanone (481)	246650[a]	271529[b]	304263[b]
unknown $C_6H_{12}O_2$ (730)	2504[a]	258669[b]	318542[b]
unknown $C_5H_{10}O_2$ (740)	570[a]	3222[b]	5798[c]
(Z)-β-Ocimene (1384)	96547[a]	25191[b]	66795[a]
Ocimene (1425)	30608[a]	7395[b]	23594[a]
Heptanoic acid (1506)	54206[a]	7147[b]	83194[b]
unknown $C_7H_6O_2$ (2568)	4583[a]	9280[b]	18219[c]
3-isobutyl-2-Methoxypyrazine (2846)	256075[a]	419161[b]	437542[b]
unknown $C_6H_8O_2$ (4453)	43825[a]	36855[a]	37338[a]
unknown $C_{15}H_{24}$ (5851)	20220[a]	19903[a]	24016[b]

Similar letters between harvests indicate that the overall intensity pattern is not significantly different (paired T-test, $p<0.05$). Metabolite intensities (compared between harvests) are shown in color scale: the darker the green color, the higher the intensity. Values between brackets indicate the mass peak used to determine peak intensity. Metabolites from the green (top) and blue (bottom) ellipses (Fig. 3C) are separated by the horizontal line

Relation between biochemical compounds and attributes

We also followed another approach to find metabolites related to flavor and/or responsible for the observed flavor differences between harvests. A Random Forest (Breiman 2001) regression approach was used on individual harvests and on the three harvests together to relate each sensory attribute to the volatiles and non-volatiles and to determine importance of the individual compounds.

Metabolites with significant ($p<0.05$) contribution to the attributes aroma, fruity/apple flavor, sweetness and sourness are summarized in Table 4. Importance of

the compounds in the attribute models is expressed by the percentage increase in mean square error (MSE) after permutation and where possible the known flavor description of found compounds is given. As physical properties of the fruits other than dry matter content were not measured, results of the texture related attributes juiciness, toughness, crunchiness and stickiness are not included.

Table 4 Compounds with significant contribution in the attribute prediction of the three individual harvests

Attribute	Compound[a]	MSE 1[b]	Corr[c]	MSE 2	Corr	MSE 3	Corr	MSE 1-3	Corr	Flavor description	Reference
Aroma	(E)-2-Hexen-1-ol (382)	31.1	0.60					5.7	0.42	Almond, fruit, spicy	Luning et al. 1994a
	Fructose	15.9	0.60			19.5	0.64	26.4	0.54	Sweet	
	Neopentane (65)	14.2	0.45					2.2	0.31	-	
	Brix	3.4	0.49			24.1	0.62	7.7	0.47	-	
	(Z)-Linalooloxide (1708)	2.9	0.40							Floral, green bell pepper	Luning et al. 1994a
Fruity	(E)-2-Hexen-1-ol (382)	5.1	0.52							Almond, fruit, spicy	Luning et al. 1994a
	Fructose	4.6	0.56			2.8	0.48	13.7	0.39	Sweet	
	Neopentane (65)	3.9	0.35							-	
	Glucose	2.9	0.52			6.9	0.50	3.3	0.36	Sweet	
	p-Menth-1-en-9-al (3632)	2.2	0.52			5.8	0.13			Spicy, herbal	Good scents company[d]
	Unknown $C_{10}H_{16}O$ (4633)					35.1	-0.10			-	
	3-Hepten-2-one (608)					8.2	-0.52	3.3	-0.33	Creamy, coconut, cheesy	Mosciano et al.1993
	Valencene (5995)					3.1	-0.18			Sweet, fresh, citrus, grapefruit	Good scents company
Sweet	(E)-2-Hexen-1-ol (382)	71.4	0.54			2.4	0.33	26.0	0.34	Almond, fruit, spicy	Luning et al. 1994a
	Glucose	26.6	0.61			7.0	0.49	11.3	0.48	Sweet	
	Neopentane (65)	25.9	0.41					9.8	0.25	-	
	Fructose	16.5	0.65	3.3	0.31	10.0	0.56	30.7	0.53	Sweet	
	(Z)-β-Ocimene (1384)	6.9	0.22			2.9	-0.31	5.3	0.06	Sweet, herbal	Good scents company
	unknown $C_{15}H_{24}$ (5851)	3.6	-0.25	8.4	-0.54	3.8	-0.45	9.6	-0.37	-	
	unknown $C_{10}H_{16}$ (2215)	3.6	0.34							-	

Attribute	Compound[a]	MSE 1[b]	Corr[c]	MSE 2	Corr	MSE 3	Corr	MSE 1-3	Corr	Flavor description	Reference
Sweet	unknown $C_6H_8O_2$ (4453)	3.3	-0.29					9.4	-0.24	-	
	(Z)-2-Penten-1-ol (113)	3.2	-0.10					3.8	-0.08	rubber, plastic, green	Flavornet [e]
	(E)-Geranylacetone (6481)	2.7	0.35							Floral, fruity, pear, apple and banana with tropical nuances	Mosciano 1998
	p-Menth-1-en-9-al (3632)			5.5	0.49					Spicy, herbal	Good scents company
	1-methyl-1,4-Cyclohexadiene (3801)			5.0	0.37	2.0	0.19	5.0	0.31	fruity	Good scents company
	2-methyl-Butanoic acid hexyl ester (3037)			3.5	0.43					fruity	Rodriguez-Burruezo et al. 2010
	3-Hepten-2-one (608)					5.7	-0.53	5.8	-0.38	Creamy, coconut, cheesy	Mosciano et al. 1993
	unknown $C_{13}H_{18}$ (6373)					4.0	-0.45	2.9	-0.28	-	
	unknown $C_{10}H_{14}O$ (1837)					3.9	-0.43			-	
	Pentanoic acid heptyl ester (4594)					2.6	-0.44			Fruity	Good scents company
Sour	unknown $C_6H_8O_2$ (4453)	2.7	0.38					5.4	0.28	-	
	Caryophyllene (7584)			4.1	0.36					Spicy, woody with citrus backgr	Good scents company
	1-Hexanol (424)			2.8	-0.47					Fruity, green bell pepper, herbal	Luning et al. 1994a
	unknown $C_{12}H_{14}$ (4187)			2.8	-0.36					-	
	unknown C_5H_{10} (100)			2.6	-0.35					-	
	2-methyl-Butanoic acid hexyl ester (3037)			2.2	-0.31					fruity	Rodriguez-Burruezo et al. 2010
	p-Menth-1-en-9-al (3632)			2.1	-0.35					Spicy, herbal	Good scents company
	Acetic acid butyl ester (211)			2.0	-0.31					Sweet, ripe banana, tropical with green nuances	Mosciano 1999

[a] Numbers between brackets refer to the compound number in Online Resource 1.[b] Percentage increase in mean square error after permutation. [c] Pearson's correlation coefficient based on log transformed metabolic data. [d] http://www.thegoodscentscompany.com , [e] http://www.flavornet.org/d_odors.html

Several common attribute predictors for aroma, fruity/apple and sweetness were found between the first and third harvest, like (E)-2-hexen-1-ol, glucose, fructose, 1-methyl-1,4-cyclohexadiene and (Z)-β-ocimene for sweetness. In the second harvest however, mainly unique compounds were found, especially for the attribute sourness. Compounds which were predictive in more than one harvest always had the same correlation direction between harvests, with the only exception of (Z)-β-ocimene versus sweetness, which had respectively a positive and negative correlation in the first and third harvest. The flavors of all these common compounds matched with the attributes aroma, fruity/apple and sweetness, as they had sweet, spicy, almond and/or fruity descriptions. The average intensity patterns and concentrations of the compounds with predictive value in the analysis with the three harvests together and in one or more individual harvests, are listed in Table 5. Significant differences between harvests are indicated. An unknown $C_6H_8O_2$ metabolite and caryophyllene, with a spicy pepper like, woody with a citrus background description, were the only compounds with a positive contribution to the attribute sourness (harvest 1). The unknown $C_6H_8O_2$ metabolite was at the same time found to be negatively correlated to sweetness. All other predictors for sourness were negatively correlated (harvest 2), including 2-methyl-butanoic acid hexyl ester and p-menth-1-en-9-al, which at the same time were positively correlated to sweetness.

Table 5 Compounds with predictive value in multiple harvests

Compound[a]	Attribute[b]	Harvest 1	Harvest 2	Harvest 3
(E)-2-Hexen-1-ol (382)	a/f/s	212136^{a}	190917^{a}	149251^{b}
Neopentane (65)	a/f/s	15694^{a}	14148^{b}	13878^{ab}
p-Menth-1-en-9-al (3632)	f/s	27059^{a}	34619^{b}	27304^{a}
3-Hepten-2-one (608)	f	20621^{ab}	18989^{a}	23443^{b}
(Z)-β-Ocimene (1384)	s	96547^{a}	25191^{b}	66795^{a}
unknown $C_{15}H_{24}$ (5851)	s	20220^{a}	19903^{a}	24016^{b}
(Z)-2-Penten-1-ol (113)	s	10105^{a}	9104^{a}	9278^{a}
1-methyl-1,4-Cyclohexadiene (3801)	s	4279^{a}	5229^{b}	5256^{b}
unknown $C_{13}H_{18}$ (6373)	s	16574^{ab}	16791^{a}	15672^{b}
unknown $C_6H_8O_2$ (4453)	s/so	43825^{a}	36855^{a}	37338^{a}
Glucose	f/s	2.91^{a}	3.08^{b}	3.12^{b}
Fructose	a/f/s	2.95^{a}	3.06^{b}	3.17^{c}

[a]Average mass peak intensity patterns of volatile compounds and concentrations of fructose and glucose (g/100 g fresh weight). Values between brackets indicate the mass peak used to determine peak intensity. Metabolite intensities (compared between harvests) are shown in color scale: the darker the green color, the higher the intensity. [b]Involved attributes aroma (a), fruity/apple (f), sweetness (s) and sourness (so). Similar letters between harvests indicate that the overall intensity pattern is not significantly different (paired T-test, $p<0.05$)

For the less contrasting attributes grassiness, green bean, carrot, perfume, petrochemical and musty flavor (Figure 1D), the only interesting finding was 2-pentyl-furan with a significant contribution to the prediction of green bean flavor (correlation 0.43) and which flavor is described as green bean/butter like (http://www.flavornet.org/d_odors.html).

Attribute prediction

The constructed Random Forest models were also used for prediction of attributes, by calculation of attribute values from genotypes which were not used to fit the model, either from the same harvest(s) or from another harvest. The prediction performance of the models is expressed by the prediction R^2, which reflects the percentage of explained variance from out-of-bag samples and is therefore not the same as the goodness-of-fit as used in linear regression.

Within and between harvest predictions of the eight most contrasting attributes are summarized in Table 6. All attributes could be significantly predicted within harvest 1, whereas in the second and third harvest, respectively aroma intensity plus fruity/apple flavor and sourness were below the significance threshold ($p<0.05$). In general the predictions of the texture related attributes, juiciness, toughness, crunchiness and stickiness of the skin, were rather good and stable within and between the individual harvests, with moderate to high heritabilities (Table 2). The within-harvest predictions of the more intrinsic flavor related attributes sweetness, sourness, aroma intensity and fruity/apple flavor were however more variable between the harvests. To demonstrate this variability with the attribute sweetness as example, in Figure 4 the direct relation between sweetness and fructose in the three individual harvests is indicated. The scatter plots show that the explained variances (goodness-of-fit R^2 values) differ significantly between harvests (43%, 9% and 32%), which is directly reflected in altered prediction R^2 values (54.6%, 23.9% and 29.4%). Moreover, Figure 4 shows that the correlation between sweetness and fructose is quite dependent on the genotypes used, as removal of the extreme genotypes PBC1405 and Piquillo (Fig. 1C) has a dramatic effect on explained variance. Especially in the second harvest, where the R^2 increases more than three times (9% vs 31%) after removal of the two genotypes.

As could be expected, the predictions between harvests were generally somewhat lower than the within-harvest predictions (Table 6). The second harvest, however, reacted differently from the first and third harvest (as seen in Table 4), since the predictions of the intrinsic flavor related attributes sweetness, sourness, aroma and fruity/apple were better using the model from another harvest than the model from harvest two itself.

Table 6 Within and between harvest Random Forest prediction R^2 values of sensory attributes

Model	harv 1	harv 2	harv 3	harv 1-3	harvest 1		harvest 2		harvest 3	
Predict	harv 1	harv 2	harv 3	harv 1-3	harv 2	harv 3	harv 1	harv 3	harv 1	harv 2
Juicy	71.1	70.0	52.8	67.7	63.2	69.4	72.8	63.1	50.4	38.3
Tough	65.1	51.3	58.3	69.5	69.0	66.9	62.7	65.5	50.9	60.1
Crunchy	34.1	29.9	24.6	38.3	19.4	28.2	3.9 ns	37.8	4.0 ns	31.5
Sticky	54.3	50.6	46.5	60.3	54.5	53.0	48.6	54.4	42.5	47.4
Sweet	54.6	23.9	29.4	55.0	37.7	26.9	32.3	26.6	46.3	37.1
Aroma	42.7	5.2 ns	22.0	33.3	31.8	31.3	18.0	10.1 ns	22.0	10.2 ns
Sour	14.5	12.4	5.4 ns	25.2	15.0	28.0	25.5	15.6	27.4	4.5 ns
Fruity	31.8	0.3 ns	27.1	26.2	1.3 ns	2.0 ns	1.6 ns	3.1 ns	21.8	12.2
Average	46.0	30.4	33.3	46.3	36.5	38.2	33.2	34.5	33.2	30.2

Non significant (ns) predictions are indicated

Sweetness-sourness interaction

From all within-harvest attribute predictions in Table 6, the prediction R^2 values of sourness were lowest. Previously we suggested already that this might be caused through masking of sour related metabolites by other volatile or non-volatile compounds (Eggink et al. 2012a). Still the best prediction of sourness was made in the first harvest, with an R^2 value of 14.5%, just above the significance threshold. The only compound with a significant contribution to sourness in this harvest however, was an unknown $C_6H_8O_2$ volatile compound (Table 4), whereas no organic acids were found. From the organic acids, citrate showed the highest correlation with sourness, though only with an explained variance of 12.5% and after removal of PBC1405 and Piquillo of 5.7% (Fig. 5A). Looking for interactors with citrate a strong positive correlation between fructose and citrate was found (Fig. 5B), meaning that when fructose concentrations are high, citrate concentrations are generally high as well. At the same time, it was demonstrated that fructose concentration is positively correlated with sweetness (Fig. 4) and that the attributes sweetness and sourness are negatively correlated (Fig. 1D). Subsequently, plotting of sweetness and sourness (Fig. 5C) made clear that genotypes with both high sweetness and high sourness scores do not exist (indicated by the green triangle), while genotypes with both high citrate and high fructose concentrations do exist (Fig. 5B). It could be concluded therefore that sweetness, and thus indirectly fructose and/or other sugars at higher concentration, can mask sourness and the effect of sourness related metabolites like citrate. On the other hand, citrate interacts with sweetness as well, which can be visualized through

the genotypes PBC1405 and Piquillo that are relatively high in fructose, but score low in sweetness (Fig. 4) since they are very high in citrate concentration (515-634 mg/100 g fresh weight; Fig. 5A).

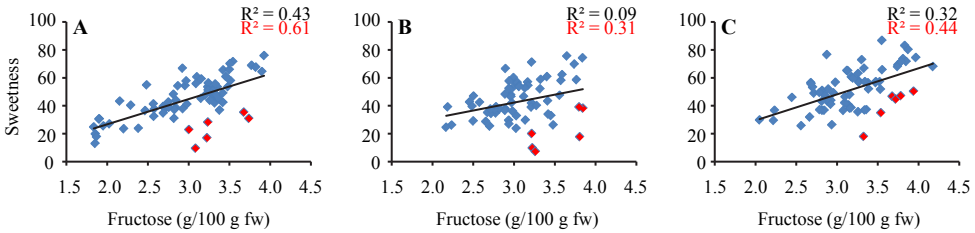


Figure 4. Sweetness-fructose scatter plots of harvest 1 (**A**), 2 (**B**) and 3 (**C**). R^2 values are indicated including (black) and excluding (red) genotypes PBC1405 and Piquillo (red dots)

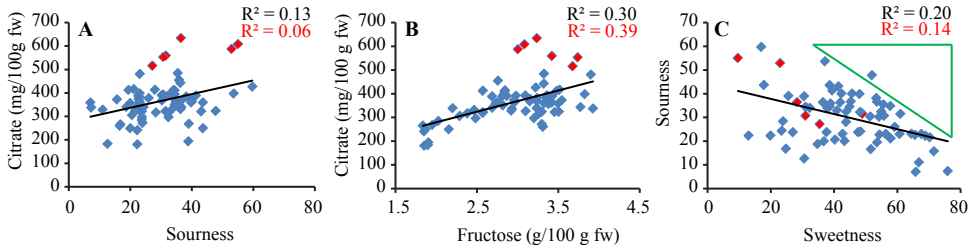


Figure 5. Sourness masking example with harvest 1 data. R^2 values are indicated including (black) and excluding (red) genotypes PBC1405 and Piquillo (red dots). The green triangle (**C**) indicates that genotypes with both high sweetness and high sourness scores are absent

Discussion

Prediction of attributes within and between harvests works well

We used a Random Forest regression approach, making use of the intensity patterns from all metabolites, for prediction of the individual flavor attributes within and between harvests. In this approach we used both a double ten-fold cross-validation, to quantify the quality of the predictions, and made predictions of samples that were not used to make the models. This combination makes the approach more powerful for prediction and identification of relevant predictors than other regression approaches that calculate goodness-of-fit R^2 values, of which the predictive value is still not known. In general, the Random Forest predictions of the texture related attributes

(juiciness, toughness, crunchiness and stickiness of the skin) and sweetness were very good. The predictions of the attributes aroma intensity, sourness and fruity/apple were somewhat lower and more variable between harvests, especially in the second harvest. In few cases the prediction of an attribute was slightly better using the model from another harvest than from the harvest itself. This could be caused by somewhat overestimated between-harvest predictions or due to the (accidental) construction of a suboptimal harvest-attribute model. It is realized that extreme genotypes could have influenced the prediction levels, as exemplified by the change in explained variance between sweetness and fructose after leaving out the genotypes PBC1405 and Piquillo. Attributes with higher heritabilities were, in general, predicted better and more consistent over harvests.

Flavor differences between harvests are caused by a combination of factors

The structure of the PCA plots with the first two principal components based on the combined three harvests (Fig. 1) is very similar to the structure of the plots based on the data from only the first harvest (Eggink et al. 2012a), indicating that there are no major differences between harvests. Giving a closer look to the scores of the individual sensory attributes per harvest revealed, however, modest, though significant harvest differences for all attributes. A similar analysis for the metabolites resulted in a list of twenty volatiles, which were most contrasting between the harvests (Table 3). Some of these volatiles could actually be involved in pepper flavor determination, like 3-isobutyl-2-methoxypyrazine, which is commonly described in pepper sniffing port analyses as characteristic (green) bell pepper aroma (Luning et al. 1994a, Rodriguez-Burruezo et al. 2010), hexanoic acid with a sweaty aroma (Brewer 2009) or (Z)-β-ocimene with a sweet, herbal description (http://www.thegoodscentscompany.com). Subsequent comparison of the twenty volatiles with all compounds that could be directly related to the attributes (Table 4), revealed however only four compounds in common: acetic acid butyl ester, (Z)-β-ocimene, an unknown $C_6H_8O_2$ volatile (compound 4453) and an unknown $C_{15}H_{24}$ volatile (compound 5851). This suggests, therefore that the other sixteen volatiles, although significantly different between harvests, do not seem to play a major role in flavor (differences between harvests). An explanation for this could be, that the harvest effects of the twenty volatiles were only noticed in PC7 versus PC3, which together represent only 7.5% of the metabolic variation. Although principal components analysis is a very efficient method to quickly visualize and simplify complex data, in this specific case it proved not as suitable as our used regression approach for identification of the responsible metabolites for the observed flavor differences

between harvests.

The flavor differences seem more likely to be caused by the observed concentration differences of the key-metabolites which were found in multiple harvests to have predictive value for the attributes aroma, fruity/apple and sweetness; (E)-2-hexen-1-ol, neopentane, p-menth-1-en-9-al, 3-hepten-2-one, (Z)-β-ocimene, (Z)-2-penten-1-ol, 1-methyl-1,4-cyclohexadiene, glucose, fructose and three unknown volatiles (Table 5), in combination and/or in interaction with the compounds that were found to be contributing to flavor attributes in individual harvests. Some studies have been performed to understand such interactions between metabolites and their effect on flavor perception in tomato (e.g. Tandon et al. 2000 & 2003, Baldwin et al. 2008), but clearly needs more attention in pepper. The relevance of the found predictors in our study was confirmed by the correspondence of the (known) flavor description of these compounds with the predicted flavor attribute(s). The contribution of (Z)-β-ocimene specifically to sweetness, is however still not completely clear, since it had a positive correlation with sweetness in the first harvest, but a negative correlation in harvest three.

As physiological maturity of the fruits at harvesting, post-harvest treatment and biochemical and sensory analyses remained the same in all three harvests, the observed differences in biochemical composition seem to be caused by differences in environmental conditions during cultivation. On the other hand we are also aware that some of the found flavor differences between harvests could also be partially caused by slight changes in the evaluations of the sensory expert panel, that evaluated the fruits on three different moments divided over a period of more than three months.

Sourness can be masked by sweetness

The described sweetness-sourness interaction explained why we did not find organic acids contributing to the prediction of sourness. Such masking of sourness and sourness related metabolites, like citrate, by fructose and other sugars has not been reported before in pepper.

Masking effects of sugars were, however, found in tomato spiking experiments, where addition of sugars and acids together to tomato puree gave a similar ripe and sweet enhancing effect on flavor perception as adding sugars alone, while addition of only acids moved flavor descriptions towards sour, tropical and citrus tastes (Baldwin et al. 2008). Nevertheless, in a large study with fresh tomatoes (Hageman et al. 2010) the masking effect of sugars on acids was not clear, since a correlation of 0.47 was found between citrate and sourness, while at the same time fructose and citrate were also positively correlated (0.57), like in our study. An explanation for this

difference between fresh tomato and pepper could lie in the fact that tomato fruits are composed of distinct edible tissue types including pericarp and locular gel, which are independently perceived in sensory experiments and which can considerably differ in chemical composition (e.g. Moretti et al. 1998), while the only edible part of pepper fruits is the pericarp.

Perspectives for flavor improvement by breeding

The flavor of pepper fruits is a complex trait, which is influenced by many factors, like the environmental conditions in which the fruits were grown, the interaction between many flavor related metabolites and the flavor perception and/or preferences of consumers. In this study we tried to investigate some of these aspects to understand their behavior better. We have shown that there are flavor differences between fruit sets (harvests) during the year. The differences over harvests between genotypes are, however, still larger than the variation within genotypes, as genotype repetitions cluster together over harvests (Fig. 1). Prediction of attributes from one harvest to another works well and we have found moderate to high heritabilities of the attributes. In addition, key-metabolites were identified that influence the sensory attributes. This combination of factors makes targeted improvement of flavor components (attributes) by breeding feasible. Still a complicating factor is that the effect and interaction of individual attributes on overall consumer liking is not completely clear. Verkerke and Janse, however, reported already in 1998, that a 'good tasting' pepper should be sweet and crunchy with a fruity aroma. New consumer preference studies are required to confirm this, but it suggests already that compounds like fructose, glucose, (E)-2-hexen-1-ol, neopentane and p-menth-1-en-9-al, with a positive contribution to both sweetness and fruity/apple flavor, are interesting candidates to increase the concentration of by breeding.

Acknowledgements

The authors kindly acknowledge Harry Jonker and Yvonne Birnbaum at Plant Research International for performing the SPME-GC-MS analyses. In addition we thank all people at Rijk Zwaan who took care of perfect greenhouse management.

Electronic supplementary material

The online version of this article (doi:10.1007/s10681-012-0761-6) contains supplementary material.

Chapter 5

Capturing flavors from *Capsicum baccatum* by introgression in sweet pepper

P.M. Eggink, Y. Tikunov, C. Maliepaard, J.P.W. Haanstra,
H. de Rooij, A. Vogelaar, E.W. Gutteling, G. Freymark,
A.G. Bovy, R.G.F. Visser
Submitted to Theoretical and Applied Genetics.

Abstract

The species *Capsicum baccatum* includes the most common hot peppers of the Andean cuisine, known for their rich variation in flavors and aromas. So far the *C. baccatum* genetic variation remained merely concealed for *C. annuum* breeding, due to post-fertilization genetic barriers encountered in interspecific hybridization. However, in order to exploit the potential flavor wealth of *C. baccatum* we combined interspecific crossing with embryo rescue, resulting in a multi-parent BC_2S_1 population. Volatile and non-volatile compounds, as well as some physical characters were measured in mature fruits, in combination with taste evaluation by a trained descriptive sensory panel. An enormous variation in biochemical composition and sensory attributes was found, with for almost all traits, BC_2S_1 genotypes showing individual values higher and/or lower than the most extreme parent (transgression). A population specific genetic linkage map, spanning 602.5 cM, was developed for QTL mapping of characterized traits. BC_2S_1 QTLs were validated in an experiment with near-isogenic lines (NILs), resulting in confirmed genetic effects for both physical, biochemical and sensory traits. Three findings are described in more detail: (i) A small *C. baccatum* LG3 introgression caused an extraordinary effect on flavor, resulting in significantly higher scores for the attributes aroma, flowers, spices, celery and chives. In an attempt to identify the responsible biochemical compounds few consistently up- and down-regulated metabolites were detected. (ii) Two introgressions (LG10.1 and LG1) had major effects on terpenoid content of mature fruits, affecting at least fifteen different monoterpenes. (iii) A second LG3 fragment resulted in a strong increase in Brix (total soluble solids) without negative effects on fruit size. The followed mapping strategy, the potential application of studied traits and perspectives for breeding are discussed.

Keywords: sensory evaluation, QTLs, NILs, unexpected flavors, terpenoids, Brix

Introduction

The genus *Capsicum* originates from South-America and comprises approximately 30 recognized species (Moscone et al. 2007). Five species of *Capsicum* are cultivated, including the closely related species *C. annuum*, *C. chinense* and *C. frutescens* that belong to the *C. annuum* complex. The other two domesticated species, *C. baccatum* and *C. pubescens*, are less known and still predominantly confined to Latin America. *Capsicum baccatum* includes the most common hot peppers (both fresh and dried) of the Andean countries and has been domesticated in the highlands of Peru and Bolivia.

The species is typically characterized by having flowers with a white corolla with basal green/yellow spots. *C. pubescens* is also a highland species, with thick-walled fleshy fruits, flowers with purple corolla and characteristic black seeds (Pickersgill 1997).

Although the domesticated species are of tropical origin, most *Capsicum* breeding has been carried out in temperate countries, and most has concerned *C. annuum* (Poulos 1994). Some wild species have however, been used in *C. annuum* breeding programs focusing on (mainly) disease resistance, such as introgression of tomato spotted wilt virus resistance from *C. chinense* (Black et al. 1991) or tobacco mosaic virus resistance from *C. chacoense* (Boukema 1982). The use of the species *C. baccatum* in *C. annuum* breeding programs has been very limited so far, since interspecific hybrization between both species is greatly hampered by post-fertilization genetic barriers (Yoon et al. 2006). Studies with *C. baccatum* focused, therefore mainly on variation of accessions within the species, showing great variability for fruit quality characteristics, yield, resistances and bioactive compounds (Do Rêgo et al 2009, Rodriguez-Burruezo et al. 2009, Yoon et al. 2004). Genetic analyses of traits from *C. baccatum* are however lacking, with as exception the molecular characterization of resistance to pepper fruit anthracnose derived from *C. baccatum* PI594137 in an intraspecific population (Kim et al. 2010). Thorough genetic analyses of *C. baccatum* (fruit quality) characteristics in interspecific mapping populations are still completely missing.

Here we present the biochemical, sensory, agronomical and molecular characterization of genotypes with *C. baccatum* var. *pendulum* introgressions in *C. annuum* genetic background. To facilitate the molecular mapping, a genetic map was constructed using a multi-parent BC_2 population. QTL mapping in a BC_2S_1 population followed by a validation experiment with near-isogenic lines (NILs) allowed the detection and confirmation of genetic effects for both physical, biochemical and sensory traits. We discuss the introgression of several unexpected traits and demonstrate that *Capsicum baccatum* is a valuable source for enrichment of the *C. annuum* breeding pool.

MATERIALS AND METHODS

Plant material

The *Capsicum baccatum* var. *pendulum* accession PEN45 was used as donor parent for backcrossing (BC) with three cultivated *C. annuum* blocky breeding lines (MT, SM and GNM) provided by the vegetable breeding company Rijk Zwaan. Due to

difficulties in interspecific crossing a multi-parent BC_2 population, consisting of three sub-populations, was generated for linkage map development (Fig. 1). The largest PEN45 BC_2 sub-population with the blocky parents SM and GNM in its pedigree (Fig. 1c) was chosen to study fruit characteristics in more detail. In this population 34 from the in total 54 BC_2 plants gave sufficient inbred seeds to grow BC_2S_1 lines. In 2009 the 34 BC_2S_1 lines were grown in plots of 5-9 plants with, when possible, 2 repetitions (possible for 23 BC_2S_1 lines) in a randomized block design. Plants were grown in soil in a greenhouse at Rijk Zwaan (De Lier, The Netherlands), according to Dutch pepper management conditions with 2.5 plants/m^2. Due to the generation of the material and the presence of two different breeding lines (SM and GNM) in their pedigree, the lines were still segregating for several traits. To grow the BC_2S_1 lines as uniform as possible, plants were pre-selected with a marker based on the Pun1 locus (Stewart et al. 2005) for selection of non-pungent plants and with a marker based on the CCS gene (capsanthin-capsorubin synthase; Lefebvre et al. 1998) to select non-red (i.e. yellow or orange) plants. To compensate for selection against Pun1 or CCS linked PEN45 fragments, with potentially interesting flavor characteristics, two and five BC_2S_1 lines (out of the 34 lines) were used to select plants with homozygous pungent orange fruits and homozygous non-pungent red fruits, respectively. These plants were also grown in 2 repetitions with plots of 5 plants. Genotypes SM, GNM and PEN45 were grown as controls in four repetitions. At the time of maturation of the first fruits the BC_2S_1 plots were made phenotypically more uniform by removing the most aberrant, mainly sterile, plants from the plots. In total 25 lines were uniform for orange fruit color, the other 9 lines were segregating for plants with either orange or yellow fruits. In the end 250 BC_2S_1 plants remained in 69 plots (1-6 plants), of which 160 orange, 61 yellow and 29 red fruited plants, that were used for QTL mapping.

Three BC_2S_1 plants, from different BC_2 plants, were used to develop near-isogenic lines (NILs) by one generation of backcrossing with GNM followed by two selfing steps. Each generation was genotyped with SNPs flanking the original BC_2S_1 introgressions to obtain BC_3S_2 lines with a limited number of introgressions in GNM genetic background. In 2011 23 NILs and the recurrent parent (GNM) were grown in three repetitions with 5 plants per plot in a completely randomized setup. Plants were grown under similar conditions as the BC_2S_1 lines in a greenhouse at Rijk Zwaan, however this time in autumn and on rockwool.

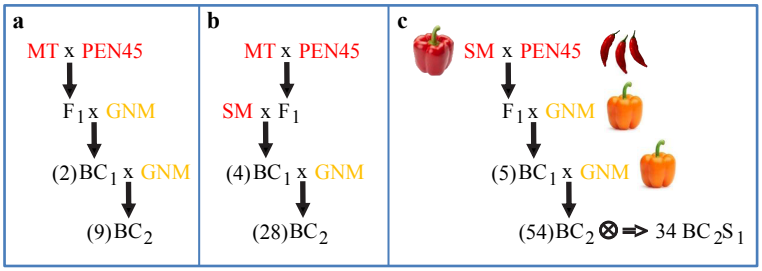


Figure 1. Three BC_2 sub-populations derived from *C. baccatum* var. *pendulum* PEN45 with *C. annuum* parents MT, SM and GNM. The number of BC-plants per generation are indicated between brackets. The different colors represent the fruit color of the parents (representative fruits indicated in c).

Trait evaluations

Ripe fruits (95-100% colored) from in general the second fruit set were used for biochemical measurements and sensory evaluation. Fruits of the BC_2S_1 plants were harvested per plot (harvest 22 May) and in case of plots segregating for plants with either orange or yellow fruits, the two colors were bulked separately. 56 BC_2S_1 plots (37 orange, 15 yellow and 4 red) gave sufficient fruits to make representative fruit samples of 5-8 fruits for sensory evaluation. In addition 32 samples were made of plots and/or individual plants that did not give enough fruits for sensory evaluation or that were pungent. In the NIL experiment (harvest 17 October), 20 NILs and GNM gave sufficient fruits and were evaluated as bulks per plot. In both experiments, fruits were stored after harvesting in a climate room at 20°C with 80% relative humidity for 3-4 days to optimize ripening. This is a standard procedure to mimic the Dutch commercial system. During the day of sensory evaluation, fruits were washed with cold tap running water, dried with a clean towel and cut (top and bottom parts were discarded and seeds were removed) in 1-2 cm pieces. Half of the fruit pieces from each sample were immediately frozen in liquid nitrogen, ground in an electric mill and stored at -80°C, while the other half was used for flavor evaluation.

A fruit description of all 250 BC_2S_1 plants and controls was made in the first week of July 2009. The shape of the fruits (conical or blocky) was recorded and average length and maximum width (cm; length1 and width1) were estimated by eye from all full grown (ripe and unripe) fruits hanging on the plant, by an experienced breeder using 0.5 cm intervals. In addition mature (orange/yellow/red) and immature (light green/dark green/pale green) fruit color were recorded. Subsequently the mature fruits were harvested and pooled per plot (76 samples excluding controls). Average weight (gram), length and width (cm; length2 and width2) were measured on 5 representative fruits. Finally, fruits were cut and roughly evaluated for odor (nasal)

intensity (scale 0: no odor – 7: a lot of odor, like PEN45) by three untrained persons. From the NILs only the average length and maximum width were estimated by eye from all full grown (ripe and unripe) fruits hanging on the plant, by an experienced breeder using 0.5 cm intervals.

Sensory analysis

The descriptive analysis took place in a sensory laboratory at Wageningen UR Greenhouse Horticulture (WUR-GH, Bleiswijk, The Netherlands). Fourteen panelists, who are part of a trained panel with broad experience in sensory evaluations of food products, including pepper, took part in the experiment. In the weeks prior to the test sessions, panelists participated in training sessions with either commercially available pepper genotypes (BC_2S_1 experiment) or with fruits from preselected NILs with divergent tastes (NIL experiment). During the training sessions, panelists agreed on fourteen attributes to describe the flavor sensation in the mouth/throat, which were the texture attributes crunchiness, stickiness of the skin, toughness and juiciness, the basic taste attributes sweetness and sourness and the retronasal flavor attributes aroma intensity, grassiness, green bean, carrot, fruity/apple, perfume, petrochemical and musty. During the training session with the preselected NILs the following five attributes were added to the list; flowers, spices (non-pungent), celery, chives and bitter, while the attribute perfume was no longer used. In both experiments the panel did not separately evaluate the odor (nasal) of the fruits.

During the BC_2S_1 experiment test sessions, each panelist (n=13) evaluated 21-22 genotypes in a randomized order, split over 2 subsequent days. On both days, 2 sessions with 5-6 genotypes and a reference (commercial orange blocky *C. annuum* hybrid) were evaluated per panelist. In this setup each sample was evaluated by 4-5 panelists. For the NIL experiment the panelists (n=12) were divided in three groups of four persons. Each group evaluated the 20 NILs of a single repetition in a randomized order (complete block design), split over 2 subsequent days. On both days, 2 sessions with 5 NILs and GNM as the reference were evaluated per panelist. In both experiments each panelist received 5 fruit pieces per sample (from multiple fruits) in a ceramic cup and they were asked to mark the intensity of the attributes on a horizontal 100-mm structured line scale on paper, resulting in a scoring between 0 and 100. The pepper fruit pieces were swallowed by the panelists. Between samples, panelists rinsed their mouth with tasteless mineral water to neutralize their palate and were also allowed to eat a small unsalted cracker before rinsing their mouth. No information was provided to the panel about the genotypes.

Metabolic profiling

Biochemical profiling of both experiments was performed as described in Eggink et al. (2012a). In short, the profiling of volatile metabolites was performed using headspace SPME-GC-MS. Derived GC-MS profiles were processed by the MetAlignTM software package (http://www.metalign.nl) for baseline correction, noise estimation and ion-wise mass spectral alignment. MSClust software tool was used to reduce the ion-wise aligned data and to extract structural information of volatile metabolites (Tikunov et al. 2012). Each compound was represented by a single selective ion fragment in the following multivariate data analysis. The compounds (number of fragment ions in a mass spectrum ≥5) were then subjected to a tentative identification using the NIST mass spectral library (http://www.nist.gov). Putative identities were assigned to compounds with a mass spectra match factor ≥ 600 and retention index deviation ≤30 . The compounds which did not meet these criteria were labelled as unknowns. Volatile compound abundance (intensity) is represented as the height of a selective mass peak of a compound detected in chromatograms by MetAlign software. Intensities which were below the detection limit in certain genotypes, obtained a random value between 250 and 500.

In both experiments the concentration of sugars (fructose, glucose and sucrose) was measured by enzymatic determination (Velterop and Vos 2001). Anion exchange chromatography was used for citric and malic acid determination based on standard protocols (Dionex Corporation, Sunnyvale, CA; http://www.dionex.com/ Application Note 143 "Determination of Organic Acids in Fruit Juices"). Sugar and acid measurements were completed by pH and total soluble solids (Brix) determination.

Genotyping and genetic linkage map construction

To genotype the 250 BC_2S_1 plants from the PEN45 BC_2 sub-population with the blocky parents SM and GNM in its pedigree (Fig. 1c), 927 SNP markers were used, of which 239 SNPs were polymorphic in PEN45 versus SM and GNM. Marker phases were set in such a way that the *C. annuum* allele was scored as 'A' and the *C. baccatum* allele as 'B'. When (fruits of) multiple plants were bulked to obtain a plot-phenotype (e.g. in case of sensory evaluations) a corresponding plot-genotype was created. For this purpose an average plot genotype value was calculated with A=0, H=0.5 and B=1, and rounded towards *C. baccatum* following the rules: average plot value <0.25=A; ≥0.25 and <0.75=H; ≥0.75=B.

In order to develop a population specific genetic linkage map, all 91 multi-parent

BC_2 individuals (Fig. 1) were screened with 412 AFLP markers and the same 927 SNPs. Markers that were polymorphic between PEN45 and the *C. annuum* parents, while not being polymorphic within MT, SM and GNM (138 AFLPs and 199 SNPs) were used for linkage map construction on the combined BC_2 population. The 54 BC_2 plants derived from crosses with SM and GNM (Fig. 1c) were used for mapping a remaining set of 34 SNPs that were segregating in the 250 BC_2S_1 plants. The combined dataset was analyzed in JoinMap 4.0 (Van Ooijen 2006) with the Independence LOD algorithm used for group construction. The regression algorithm was used to calculate marker distances within linkage groups with the Kosambi mapping function (Kosambi 1944), and using linkages with recombination frequencies smaller than 0.5 and a LOD larger than 1.0. Linkage groups were named and oriented based on initial knowledge of marker groupings on an unpublished integrated map (IntMap) of Rijk Zwaan, which corresponds with the chromosome numbering and orientation of Wu et al. (2009).

Statistical analysis

The sensory data was analysed in Genstat version 12 using a linear mixed model REML (residual maximum likelihood) analysis. For the BC_2S_1 experiment a model was used with genotype, replicate and their interaction as fixed terms. Sessions (tasting sessions) within replicate/genotype combinations and panelists within sessions were taken as random terms. For the NIL experiment a model was used with genotype as fixed term. Panelist, the panelist- genotype interaction and repetitions within panelist/genotype combinations were taken as random terms. In this experiment repetition and session were completely confounded with panelist and therefore not corrected for as a combined factor. Mean values were calculated for both experiments per genotype per replicate after a correction for session and panelist effects and removal of strong outliers (if the absolute value of a standardized residual was larger than three residual standard deviations).

QTL analysis

The Interval Mapping method within the program MapQTL 6 (Van Ooijen 2009) was used for QTL identification in the BC_2S_1 experiment. A permutation test was applied to each data set (1000 permutations) to determine the LOD (Logarithm of odds) thresholds. A genome wide (GW) LOD threshold of 2.7 was used for QTL significance ($p < 0.05$). The chromosomal locations with the highest LOD scores were considered to be the most likely positions of a QTL. Graphics were produced by MapChart software (Voorrips 2002). The NIL experiment was analyzed using the non-parametric Kruskal–Wallis test within MapQTL 6 to identify markers that

showed significant ($p < 0.05$) trait associations. The analyses in both experiments were performed with $\log_2$ transformed metabolite data.

RESULTS

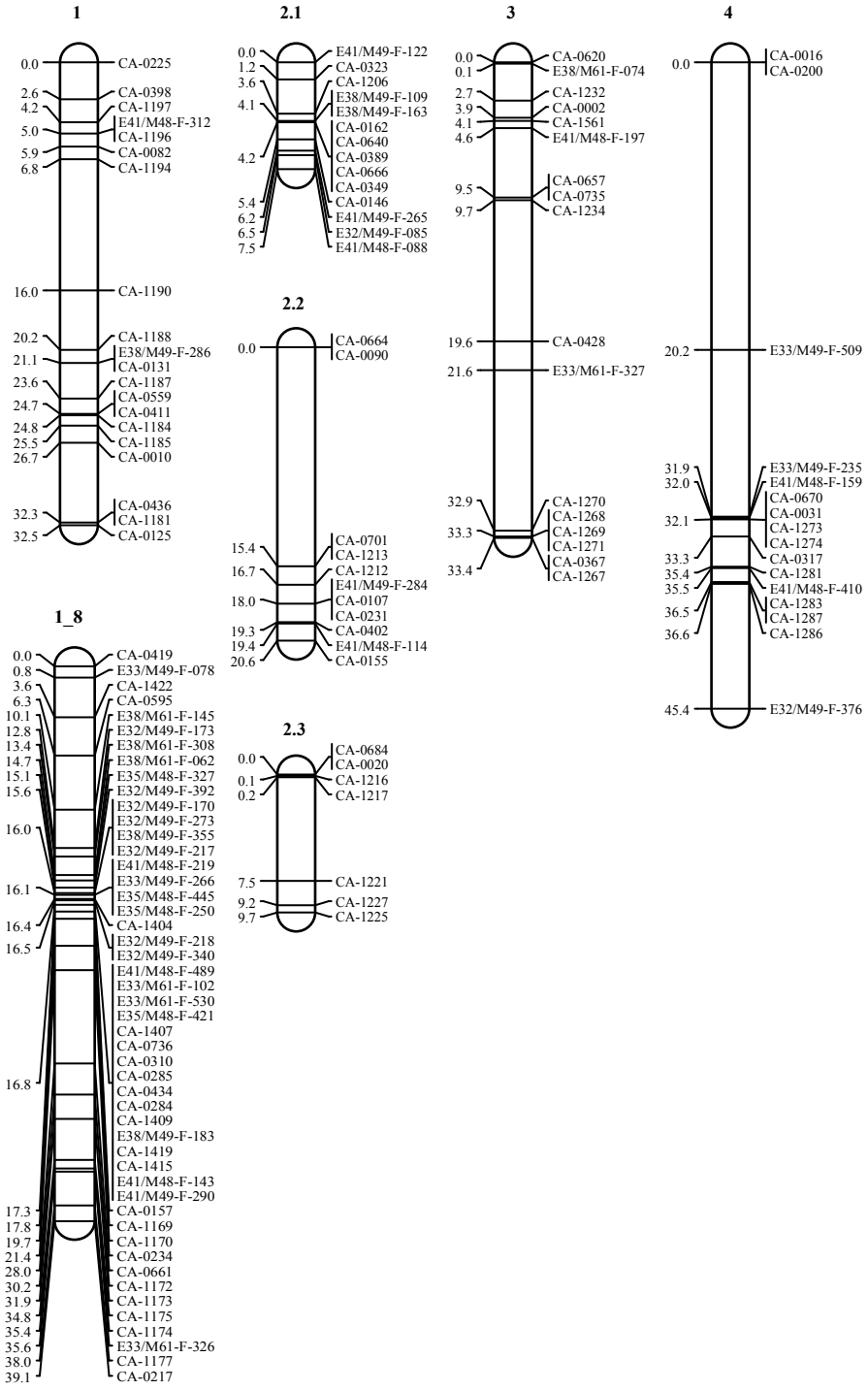
Map construction

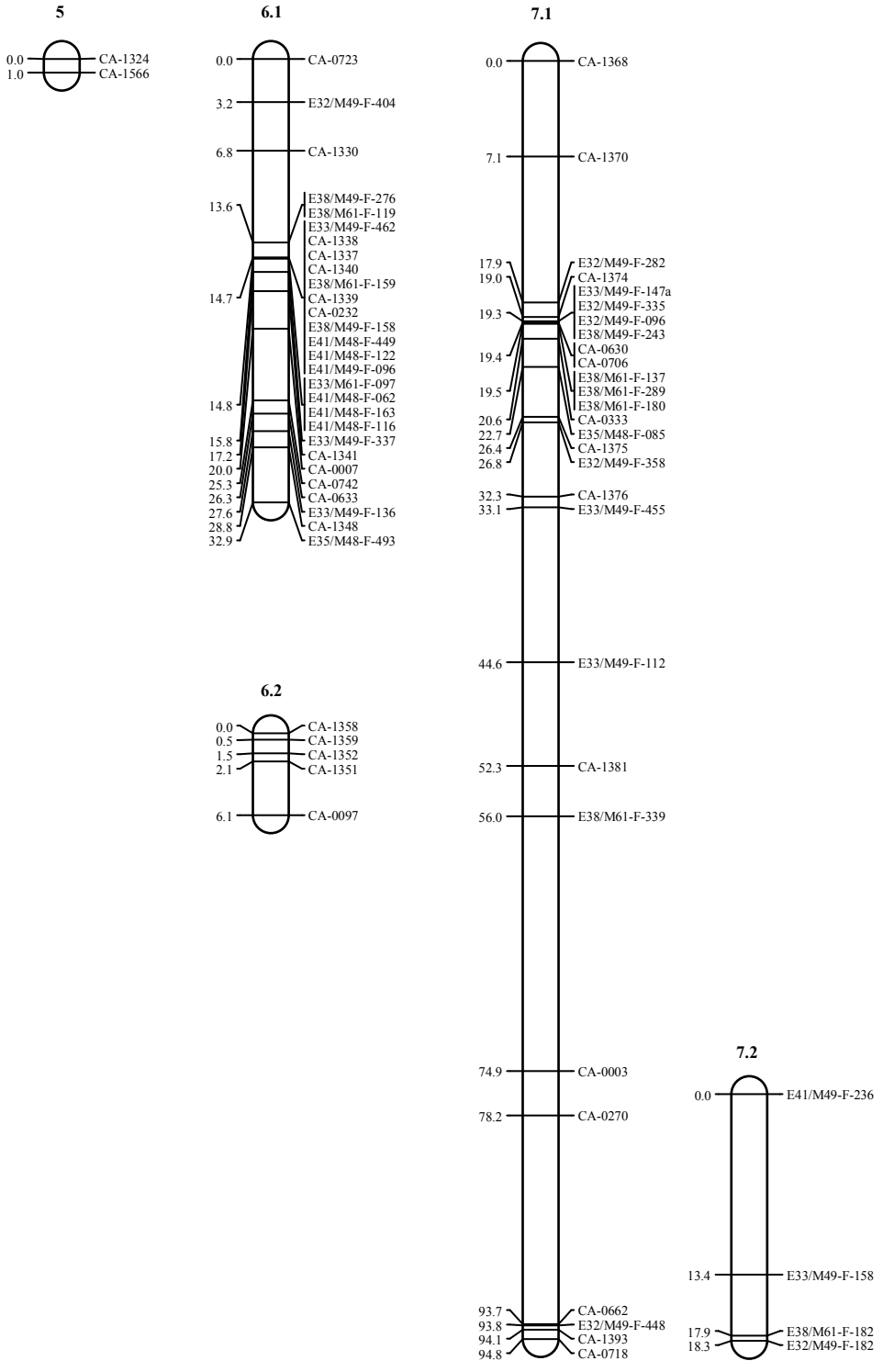
In total 412 AFLP markers and 927 SNPs were used to genotype the 91 multi-parent BC_2 individuals (Fig. 1). Almost all (366 out of 377) informative markers (138 AFLPs and 233 SNPs) were assigned to 21 linkage groups (Fig. 2), with a total map size of 602.5 cM. More than 12 linkage groups were constructed, due to absence of polymorphic markers connecting the sub-linkage groups of the 12 corresponding pepper chromosomes. Linkage groups could however be oriented and assigned to chromosomes based on an unpublished integrated map of Rijk Zwaan. This resulted in thirteen linkage groups, as for chromosome 1 and 8 an additional linkage group 1_8 was constructed, to account for the known reciprocal translocation in that region, differentiating the genome of *C. annuum* from that of other *Capsicum* species (Wu et al. 2009).

Distribution of phenotypes

In the BC_2S_1 experiment in total 222 putative volatile compounds were detected of which 22 volatiles were specific to PEN45 (i.e. under detection limit in all BC_2S_1 plants and *C.annuum* parents). Putative identities could be assigned to 178 of these, based on their library match factor and retention indices. In the NIL experiment in total 137 putative volatile compounds were detected. Identities were assigned to 96 of these. In both experiments the sucrose concentrations turned out to be under the detection limit (0.3 g/100 g fresh weight) of our enzymatic determination method.

Most of the characterized biochemical, sensory and physical traits showed continuous variation in the BC_2S_1 population, except for the traits color, shape and pungency which were scored in classes (Table 1). The small fruited PEN45 could be clearly differentiated from the *C. annuum* parents by higher sugar and especially acid concentrations, with up to three times higher malate and citrate levels. A number of selected aroma compounds, representing the major metabolic pathways and known to have an effect on pepper flavor (Table 2) differed also clearly between the parents. Due to its pungency, PEN45 was not included in the sensory evaluations and therefore its attribute scores, except for odor, are missing in Table 1. For almost all traits the BC_2S_1 plants or plots showed individual values higher and/or lower than the most extreme parent, a phenomenon known as transgression.


1
0.0 CA-0225
2.6 CA-0398
4.2 CA-1197
5.0 E41/M48-F-312
CA-1196
5.9 CA-0082
6.8 CA-1194
16.0 CA-1190
20.2 CA-1188
21.1 E38/M49-F-286
CA-0131
23.6 CA-1187
24.7 CA-0559
CA-0411
24.8 CA-1184
25.5 CA-1185
26.7 CA-0010
32.3 CA-0436
CA-1181
32.5 CA-0125
2.1
0.0 E41/M49-F-122
1.2 CA-0323
3.6 CA-1206
4.1 E38/M49-F-109
E38/M49-F-163
CA-0162
CA-0640
4.2 CA-0389
CA-0666
CA-0349
5.4 CA-0146
6.2 E41/M49-F-265
6.5 E32/M49-F-085
7.5 E41/M48-F-088
2.2
0.0 CA-0664
CA-0090
15.4 CA-0701
CA-1213
16.7 CA-1212
E41/M49-F-284
18.0 CA-0107
CA-0231
19.3 CA-0402
19.4 E41/M48-F-114
20.6 CA-0155
2.3
0.0 CA-0684
CA-0020
0.1 CA-1216
0.2 CA-1217
7.5 CA-1221
9.2 CA-1227
9.7 CA-1225
3
0.0 CA-0620
0.1 E38/M61-F-074
2.7 CA-1232
3.9 CA-0002
4.1 CA-1561
4.6 E41/M48-F-197
9.5 CA-0657
CA-0735
9.7 CA-1234
19.6 CA-0428
21.6 E33/M61-F-327
32.9 CA-1270
CA-1268
33.3 CA-1269
CA-1271
CA-0367
33.4 CA-1267
4
0.0 CA-0016
CA-0200
20.2 E33/M49-F-509
31.9 E33/M49-F-235
32.0 E41/M48-F-159
CA-0670
32.1 CA-0031
CA-1273
CA-1274
33.3 CA-0317
35.4 CA-1281
35.5 E41/M48-F-410
CA-1283
36.5 CA-1287
36.6 CA-1286
45.4 E32/M49-F-376
1_8
0.0 CA-0419
0.8 E33/M49-F-078
3.6 CA-1422
6.3 CA-0595
10.1 E38/M61-F-145
12.8 E32/M49-F-173
13.4 E38/M61-F-308
14.7 E38/M61-F-062
15.1 E35/M48-F-327
15.6 E32/M49-F-392
E32/M49-F-170
16.0 E32/M49-F-273
E38/M49-F-355
E32/M49-F-217
E41/M48-F-219
16.1 E33/M49-F-266
E35/M48-F-445
E35/M48-F-250
16.4 CA-1404
16.5 E32/M49-F-218
E32/M49-F-340
E41/M48-F-489
E33/M61-F-102
E33/M61-F-530
E35/M48-F-421
CA-1407
CA-0736
CA-0310
16.8 CA-0285
CA-0434
CA-0284
CA-1409
E38/M49-F-183
CA-1419
CA-1415
E41/M48-F-143
E41/M49-F-290
17.3 CA-0157
17.8 CA-1169
19.7 CA-1170
21.4 CA-0234
28.0 CA-0661
30.2 CA-1172
31.9 CA-1173
34.8 CA-1175
35.4 CA-1174
35.6 E33/M61-F-326
38.0 CA-1177
39.1 CA-0217


5
0.0 CA-1324
1.0 CA-1566
6.1
0.0 CA-0723
3.2 E32/M49-F-404
6.8 CA-1330
13.6 E38/M49-F-276
E38/M61-F-119
E33/M49-F-462
CA-1338
CA-1337
CA-1340
E38/M61-F-159
14.7 CA-1339
CA-0232
E38/M49-F-158
E41/M48-F-449
E41/M48-F-122
E41/M49-F-096
E33/M61-F-097
14.8 E41/M48-F-062
E41/M48-F-163
E41/M48-F-116
15.8 E33/M49-F-337
17.2 CA-1341
20.0 CA-0007
25.3 CA-0742
26.3 CA-0633
27.6 E33/M49-F-136
28.8 CA-1348
32.9 E35/M48-F-493
6.2
0.0 CA-1358
0.5 CA-1359
1.5 CA-1352
2.1 CA-1351
6.1 CA-0097
7.1
0.0 CA-1368
7.1 CA-1370
17.9 E32/M49-F-282
19.0 CA-1374
E33/M49-F-147a
19.3 E32/M49-F-335
E32/M49-F-096
E38/M49-F-243
19.4 CA-0630
CA-0706
E38/M61-F-137
19.5 E38/M61-F-289
E38/M61-F-180
20.6 CA-0333
22.7 E35/M48-F-085
26.4 CA-1375
26.8 E32/M49-F-358
32.3 CA-1376
33.1 E33/M49-F-455
44.6 E33/M49-F-112
52.3 CA-1381
56.0 E38/M61-F-339
74.9 CA-0003
78.2 CA-0270
93.7 CA-0662
93.8 E32/M49-F-448
94.1 CA-1393
94.8 CA-0718
7.2
0.0 E41/M49-F-236
13.4 E33/M49-F-158
17.9 E38/M61-F-182
18.3 E32/M49-F-182

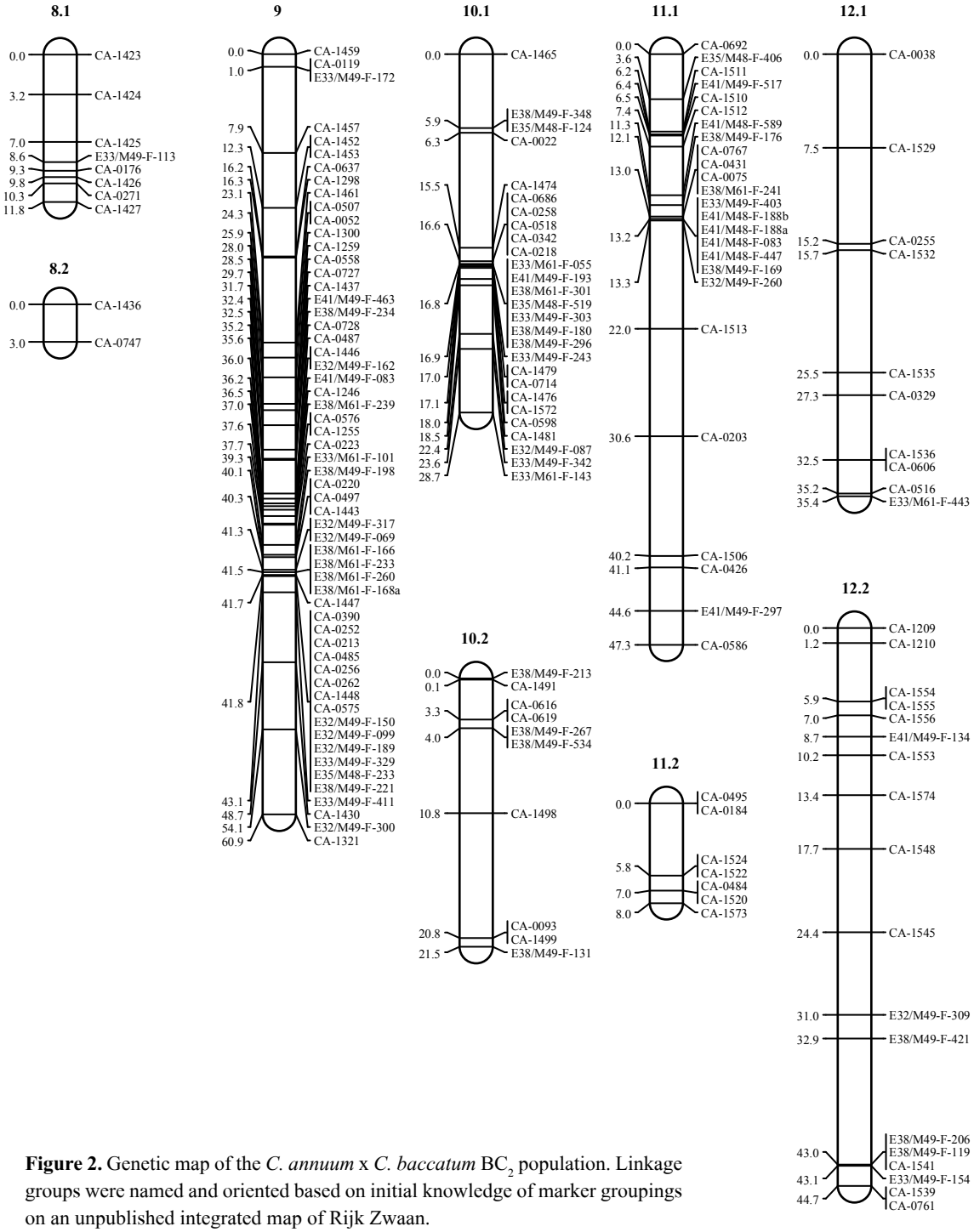


Figure 2. Genetic map of the *C. annuum* x *C. baccatum* BC_2 population. Linkage groups were named and oriented based on initial knowledge of marker groupings on an unpublished integrated map of Rijk Zwaan.

Based on the description of the sensory panel and the biochemical measurements three BC_2S_1 plants with fruits that had either an extraordinary flavor resulting from high sweetness, sourness and/or odor scores or a high sugar and acid concentration were chosen for further analysis. These three plants, originating from different BC_2 plants, were used to develop near-isogenic lines (NILs), by one generation of backcrossing with GNM followed by two selfing steps. Each generation was genotyped with SNPs flanking the original BC_2S_1 introgressions to obtain BC_3S_2 lines with one up to four homozygous, and in most cases also one or two heterozygous introgressions in GNM genetic background (Fig. 3). NILs were evaluated in comparison to the recurrent parent and again for almost all traits, NILs were found with higher as well as lower traits scores than GNM (Table 1).

QTL analysis in BC_2S_1 population and validation via NILs

The 250 BC_2S_1 plants from the PEN45 BC_2 sub-population having the blocky parents SM and GNM in its pedigree (Fig. 1c), were genotyped with 239 SNPs that were polymorphic in PEN45 versus SM and GNM. Interval mapping, with separate sessions for sensory attributes (14 attributes, 56 plots), metabolites (200 volatiles and 6 non-volatiles, 88 plots/plants) and several physical fruit characters plus odor (either on 250 plants or 76 plots), allowed identification of QTLs within all trait classes (Table 3).The traits pungency and red color were used to validate the quality of our QTL mapping results, since both traits have been mapped in other mapping populations previously. CCS (red color) was mapped by Thorup et al. (2000) on chromosome 6 around 80-100 cM. Blum et al. (2002) mapped the single dominant gene *C* (former name of Pun1), required for capsaicin synthesis, to chr. 2 at 50.1 cM (FA03 map; http://solgenomics.net). Although in our population only three out of the 250 BC_2S_1 plants were pungent, we were able to map the trait pungency with a perfect fit (LOD 99.9; Table 3) to LG2.2 at 15.4 cM, which corresponds to chr. 2, 65.4 cM on our integrated *C. annuum* map (IntMap; unpublished data). Also for red color a QTL was found in the expected CCS region on chr. 6, *i.e.* marker CA-0097 on LG6.2 at 6.1 cM (Table 3), corresponding to chr. 6, 95.4 cM (IntMap). The percentage of explained variance of the latter QTL (13.5%) was, however, surprisingly low knowing that a single dominant gene is involved. Taking a closer look at the region around CCS on the IntMap, we noticed that all markers in this region were not used for the population specific PEN45 BC_2 map, since these markers were polymorphic within the *C. annuum* parents SM (red fruits) and GNM (orange fruits). Due to the multi-parent nature of the BC_2, during map construction we thus automatically selected against markers linked to fruit color. In a subsequent single marker analysis with polymorphic markers within the *C. annuum* parents included, marker CA-0081(chr. 6, 114.7 cM; IntMap) was found with a perfect correlation to red color (data not shown).

Table 1 Variation of measured traits in the parents, BC_2S_1 population and NILs

		Parents			BC_2S_1 population[1]					NILs				
Class	**Trait**	**GNM**	**SM**	**PEN45**	**Mean**	**SD**	**Min**	**Max**	**#gen**	**GNM**	**Mean**	**SD**	**Min**	**Max**
Physical	Color ripe[2]	orange	red	red	0.12	0.32	0.0	1.0	250	orange	0.0	0.0	0.0	0.0
characters	Color unripe[3]	dg	dg	pale	0.77	0.33	0.0	1.0	250	na	na	na	na	na
	Length1 (cm)	8.1	8.3	7.0	7.5	1.3	5.0	10.0	250	8.0	7.4	1.1	4.5	9.0
	Width1 (cm)	8.5	8.3	2.0	5.7	1.5	3.0	9.0	250	8.0	7.1	1.1	4.5	8.0
	Shape[4]	block	block	conical	0.74	0.25	0.5	1.0	250	na	na	na	na	na
	Pungency[5]	nonpun	nonpun	pun	0.012	0.109	0.0	1.0	250	nonpun	0.0	0.0	0.0	0.0
	Length2 (cm)	8.6	7.6	7.4	7.5	1.3	5.0	11.8	76	na	na	na	na	na
	Width2 (cm)	8.7	8.1	2.1	6.1	1.3	4.0	8.6	76	na	na	na	na	na
	Weight (g)	209.5	188.0	8.0	92.5	46.3	32.0	224.0	76	na	na	na	na	na
Biochemical	pH	5.0	5.1	5.0	5.0	0.1	4.8	5.3	88	na	na	na	na	na
composition	Brix	8.3	8.3	11.0	10.6	2.2	6.8	17.9	88	7.2	8.1	0.7	6.9	10.2
	Glucose[6]	2.8	3.2	4.1	3.7	0.8	1.9	6.4	88	2.4	2.9	0.3	2.4	3.9
	Fructose[6]	3.1	3.0	3.3	3.8	0.9	2.2	6.7	88	2.6	3.1	0.3	2.7	4.0
	Malate[7]	21.9	24.4	79.9	37.4	16.5	14.1	115.4	88	23.5	24.3	11.0	13.0	77.3
	Citrate[7]	364.8	363.9	1037.7	450.9	121.6	210.0	820.0	88	303.6	333.9	64.2	230.0	605.6
	Hexanal	273101	515522	2967280	832246	605386	128780	2728550	88	207297	232354	103477	53200	458228
	3-Hepten-2-one	136745	101361	7956	75650	46170	10553	224728	88	22429	22526	10495	3970	51327
	Linalooloxide	17585	15577	16344	127721	132918	7904	642158	88	18537	48102	71724	4047	265487
	p-Menth-1-en-9-al	30944	33239	68804	290157	291680	8368	1355670	88	44465	109913	156677	5333	502535
	PAE	5334	378	410	3321	5903	250	25618	88	na	na	na	na	na
	BAE	356	392	64326	13109	15888	252	74676	88	na	na	na	na	na
	Geranylacetone	45271	82238	47450	62310	38618	7313	259467	88	7891	10948	10059	416	52915
	b-Damascenone	84925	30845	17766	62228	45086	5689	259570	88	10579	8965	6879	400	30777
	Methoxypyrazine	243979	259603	300809	446015	318569	16194	1403980	88	205412	220665	114681	36227	455322

		Parents			BC_2S_1 population[1]					NILs				
Class	**Trait**	**GNM**	**SM**	**PEN45**	**Mean**	**SD**	**Min**	**Max**	**#gen**	**GNM**	**Mean**	**SD**	**Min**	**Max**
Sensory	Crunchy	63.9	58.1	na	59.9	7.5	42.4	79.0	56	65.8	68.0	5.9	51.7	79.2
attributes	Sticky	38.0	55.1	na	44.4	14.1	16.6	73.6	56	12.2	19.0	6.5	5.4	35.4
	Tough	23.4	28.0	na	33.6	13.7	9.1	71.6	56	14.9	19.5	5.8	8.9	34.6
	Juicy	63.7	55.5	na	46.1	10.1	21.8	72.2	56	57.9	56.4	7.1	36.5	72.3
	Sweet	34.2	28.6	na	32.9	8.7	16.8	54.3	56	40.2	46.0	6.6	26.5	59.7
	Sour	25.9	20.2	na	22.5	8.0	6.2	41.0	56	31.6	31.8	5.6	21.8	46.5
	Aroma	38.3	32.7	na	41.3	11.6	16.1	67.6	56	45.3	55.6	10.0	33.0	73.7
	Grass	5.7	10.8	na	8.1	5.9	0.0	30.0	56	9.6	9.1	6.5	0.0	30.1
	Bean	0.3	4.6	na	6.4	5.6	0.0	23.0	56	6.2	8.0	6.7	0.0	31.7
	Carrot	3.3	1.2	na	2.2	2.5	0.0	10.9	56	7.5	4.4	3.6	0.0	14.4
	Fruity	9.9	8.5	na	12.1	8.2	0.1	41.3	56	19.8	15.2	8.2	2.9	36.0
	Perfume	0.4	1.7	na	2.3	2.4	0.0	10.1	56	na	na	na	na	na
	Flowers	*	na	na	na	na	na	na	na	1.8	5.2	4.8	0.0	22.8
	Spices	*	na	na	na	na	na	na	na	3.9	9.4	7.0	0.0	33.4
	Celery	*	na	na	na	na	na	na	na	1.1	3.9	4.4	0.0	18.3
	Chives	*	na	na	na	na	na	na	na	6.0	3.3	5.2	0.0	18.6
	Petrochemical	0.7	0.1	na	2.3	3.1	0.0	14.8	56	2.7	2.9	4.2	0.0	21.2
	Musty	2.7	1.6	na	1.1	1.6	0.0	8.6	56	4.4	1.7	2.1	0.0	8.5
	Odor	1.8	1.5	7.0	1.9	1.5	0.0	6.0	76	na	na	na	na	na

[1] The BC_2S_1 population and NILs are described by mean, standard deviation (SD) and range (min-max). The number of genotypes (#gen; plots or individual plants) used for phenotyping the BC_2S_1 population is indicated. Metabolite intensities are shown, separated per experiment, in color scale from low (light red) to high intensity (dark red). For abbreviations of metabolites see Table 2. [2,3,4,5] Class scores were translated numerically in the BC_2S_1 population for the traits color ripe (red=1; orange or yellow=0), color unripe (dark green (dg)=1; light green=0.5; pale=0), shape (blocky =1; conical =0.5) and pungency (non-pungent (nonpun)=0; pungent (pun)=1). [6] g/100 g fresh weight, [7] mg/100 g fresh weight. na= not available.

Table 2 Aroma reference volatiles

Metabolite	El. comp[1]	Type	Flavor description/reason
Hexanal	$C_6H_{12}O$	Lipid	Green[2]
3-Hepten-2-one	$C_7H_{12}O$	Lipid	Negative correlation with fruity[3]
Linalooloxide	$C_{10}H_{18}O_2$	Terpenoid	Floral, green bell pepper[4]
p-Menth-1-en-9-al	$C_{10}H_{16}O$	Terpenoid	Positive correlation with fruity[3]
Pentanoic acid, hexyl ester (PAE)	$C_{11}H_{22}O_2$	Ester	Fruity, green[5]
Butanoic acid, 2-methyl-hexyl ester (BAE)	$C_{11}H_{22}O_2$	Ester	Fruity[2]
Geranylacetone	$C_{13}H_{22}O$	Carotenoid	Sweet, citrus[6]
β-damascenone	$C_{13}H_{18}O$	Carotenoid	Fruity, floral[5]
2-isobutyl-3-methoxypyrazine	$C_9H_{14}N_2O$	Pyrazine	Green pepper[4]

[1] Elemental composition. [2] Rodriguez-Burruezo et al. 2010, [3] Eggink et al. 2012b, [4] Luning et al. 1994a, [5] http://www.thegoodscentscompany.com, [6] Tandon et al. 2000

For unripe fruit color a very strong QTL (LOD 40.1, 52.8% explained variance) was found on LG10.1 for the contrast between dark and pale green fruit color. The trait was however not measured in the NILs for validation. In addition, several QTLs were found for the physical traits length, width, shape and weight, of which some length and width effects were actually confirmed in the NILs. As expected there was a big overlap in detected QTLs for the estimated (length1 and width1) and measured (length2 and width2) size QTLs. Both measurements, however, turned out to be valuable as they delivered different QTLs which could be confirmed in the NILs (Table 3).

In the BC_2S_1 population also several QTLs for biochemical compounds were found. For Brix, glucose, fructose, malate as well as citrate a significant effect was found on LG1_8, with the *C. baccatum* allele responsible for an increase in concentration. This QTL however coincided, with the main fruit size QTL, for which both width, shape and weight decrease at the *C. baccatum* allele (Table 3). Therefore, the increase in the concentration of these non-volatiles seems to be an effect of smaller fruits rather than an absolute increase in amount. In the NILs this LG1_8 QTL for non-volatiles could not be confirmed. In contrast, a QTL for malate on LG1 could actually be confirmed in the NILs. NIL47 containing this telomeric LG1 *C. baccatum* introgression (Fig. 3) even showed an almost three fold increased concentration of malate compared to the recurrent parent and NILs lacking this introgression (Table 3).

Significant effects which could be confirmed in the NILs were also found for several aroma reference compounds, with a common QTL on LG10.1 for the metabolites linalooloxide, p-menth-1-en-9-al, butanoic acid, 2-methyl-hexyl ester (BAE) and β-damascenone (Table 3). A common QTL was found as well for hexanal,

3-hepten-2-one, geranylacetone and methoxypyrazine on LG1_8. The effect of this locus could however only be confirmed in the NILs for 3-hepten-2-one.

In the BC_2S_1 population a QTL affecting the texture related attributes toughness, stickiness of the skin and juiciness was found in the same region on LG1_8, that also gave significant effects for the fruit size traits and non-volatiles. Like for the non-volatiles, this QTL could not be confirmed in the NILs (Table 3). Interestingly, a rather strong QTL (LOD 8.0, 38.7% explained variance) for odor was found on LG3 at 33.3 cM, with the *C. baccatum* allele giving a more intense odor than the *C. annuum* allele (increase of 3 points on 0-7 scale). Since in the BC_2S_1 population odor was only scored for its intensity and not further specified, in the NILs this trait was separated into multiple attributes, as described below.

Flavor effect of C. baccatum LG3 introgression

During the sensory evaluation of the fruits in the BC_2S_1 experiment it became clear that the vocabulary (i.e. predefined attributes) of the trained panel was not sufficient to cover all the flavor variation, which resulted in remarks on the evaluation sheets like ‘presence of tropical fruit flavor’ or ‘similar to papaya taste’. This was caused by the fact that the panelists participated in training sessions with only commercially available genotypes, lacking these (unexpected) flavors. For the NIL experiment the test panel was therefore also trained with fruits from preselected NILs having more extreme flavors than currently available in the Dutch commercial segment. This resulted in an expansion of the panel’s vocabulary with the attributes flowers, spices (non-pungent), celery, chives and bitter.

Analysis of the sensory data from the NIL experiment using the non-parametric Kruskal–Wallis test, showed that NILs having either a heterozygous (NIL40, 46 and 48) or homozygous (NIL37, 38 and 39) LG3 introgression containing the BC_2S_1 odor QTL (Table 3; Fig. 3) have significantly higher scores for the attributes aroma, flowers, spices, celery and chives (Table 4). This confirmed that the small *C. baccatum* introgression on LG3 from 32.9 to 33.4 cM causes an extraordinary effect on flavor. Subsequently we checked for which metabolites there are QTLs in this LG3 region in the BC_2S_1 population, which resulted in nine volatiles (Table 4). Out of these nine metabolites, four were also detected and confirmed in the NILs. The compounds 6-methyl-4-oxo-5-heptenal and unknown_8805 showed a strong increase in intensity in the presence of the *C. baccatum* allele, while (Z)-butanoic acid 3-hexenyl ester and 2-isobutyl-3-methoxypyrazine were decreased in NILs having the introgression (Table 4). The confirmed up-regulated compounds were also checked for their direct relation to odor in the BC_2S_1 population resulting in a correlation of 0.53 for 6-methyl-4-oxo-5-heptenal and 0.43 for unknown_8805.

Table 3 QTLs detected in the BC_2S_1 population and genotype means of the NILs

		BC_2S_1 population										NILs				
Class	**Trait**	**LG**	**cM**	**Marker**	**LOD**	**%EV**	**µA**	**µH**	**µB**	**add.**	**A/H/B**	**Signif[1]**	**mA**	**mH**	**mB**	**A/H/B[2]**
Physical	Pungency	2.2	15.4	CA-0701	99.9	100.0	0.0	0.5	1.0	-0.50	247/0/3	-	-	-	-	no var
characters	Color ripe	12.2	0.0	CA-1209	8.1	13.9	0.1	0.8	1.5	-0.67	242/4/1	-	-	-	-	no var
		2.1	1.2	CA-0323	8.0	13.8	0.1	0.7	1.2	-0.57	244/3/2	-	-	-	-	no var
		6.2	6.1	CA-0097	7.8	13.5	0.1	0.4	0.8	-0.35	235/9/5	-	-	-	-	no var
		12.1	0.0	CA-0038	4.6	8.3	0.1	0.4	0.6	-0.26	231/14/5	-	-	-	-	no var
	Color unripe	10.1	18.5	CA-1481	40.1	52.8	0.9	0.6	0.2	0.36	147/64/24	-	-	-	-	not meas
		4	0.0	CA-0200	2.9	5.2	0.8	0.7	0.6	0.11	170/49/24	-	-	-	-	not meas
		7.1	93.7	CA-0662	2.8	5.1	0.8	0.9	1.1	-0.16	221/16/11	-	-	-	-	not meas
	Length1	7.1	78.2	CA-0270	5.3	10.5	7.7	6.7	5.7	0.96	219/14/11	-	-	-	-	no introgr
		10.1	16.6	CA-0686	5.1	10.2	7.8	7.2	6.6	0.60	146/75/28	**0.0001**	**7.63**	**-**	**5.50**	**57/0/6**
		1	0.0	CA-0225	4.0	7.9	7.8	7.3	6.8	0.50	132/75/39	ns	7.4	-	8.0	60/0/3
		3	2.7	CA-1232	3.7	7.3	7.4	8.2	9.0	-0.81	213/29/6	ns	7.4	-	8.0	60/0/3
	Length2	10.1	16.6	CA-0686	3.6	20.5	8.1	7.0	6.0	1.00	39/32/4	**0.0001**	**7.63**	**-**	**5.50**	**57/0/6**
	Width1	1_8	19.7	CA-1170	19.8	33.9	6.4	4.8	3.2	1.62	146/96/4	**0.05**	**7.23**	**-**	**6.00**	**60/0/3**
		10.1	15.5	CA-1474	8.9	16.9	6.2	5.3	4.4	0.88	143/78/28	**0.0001**	**7.50**	**-**	**5.17**	**54/0/9**
		9	37.6	CA-0576	7.8	15.1	5.9	4.4	2.8	1.55	214/31/2	**0.05**	**7.23**	**-**	**6.00**	**60/0/3**
		8.1	0.0	CA-1423	6.7	12.9	6.0	4.6	3.2	1.38	202/42/0	-	-	-	-	no introgr
		11.1	7.4	CA-1512	6.1	11.9	6.1	5.3	4.4	0.83	158/73/0	ns	7.2	-	7.0	60/0/3
		5	1.0	CA-1566	4.4	8.8	5.9	5.2	4.5	0.74	197/36/16	**0.005**	**7.11**	**6.75**	**8.00**	**42/12/9**
		4	36.5	CA-1283	3.6	7.3	5.9	5.0	4.0	0.93	202/41/4	ns	-	-	-	no introgr
	Width2	1_8	16.8	CA-1409	8.2	40.4	7.0	6.0	5.0	0.98	27/25/24	**0.05**	**7.23**	**-**	**6.00**	**60/0/3**
		9	40.3	CA-0220	3.9	21.7	6.4	5.0	3.6	1.40	47/16/0	ns	7.2	7.0	7.3	42/9/12
		11.1	13.0	CA-0767	3.2	18.3	6.5	5.4	4.3	1.11	46/28/0	ns	7.2	7.0	7.3	42/9/12
	Shape	1_8	16.8	CA-1409	23.1	38.0	0.9	0.7	0.6	0.16	150/3/97	-	-	-	-	not meas
		11.1	7.4	CA-1512	7.2	13.7	0.8	0.7	0.5	0.15	158/73/0	-	-	-	-	not meas

		BC_2S_1 population										NILs				
Class	**Trait**	**LG**	**cM**	**Marker**	**LOD**	**%EV**	**µA**	**µH**	**µB**	**add.**	**A/H/B**	**Signif[1]**	**mA**	**mH**	**mB**	**A/H/B[2]**
		9	37.6	CA-0576	5.0	9.9	0.8	0.6	0.3	0.22	214/31/2	-	-	-	-	not meas
		5	1.0	CA-1566	3.9	7.8	0.8	0.7	0.5	0.12	197/36/16	-	-	-	-	not meas
		8.1	0.0	CA-1423	3.0	6.0	0.8	0.6	0.4	0.16	202/42/0	-	-	-	-	not meas
	Weight	1_8	16.8	CA-1409	7.8	38.9	125.4	90.7	55.9	34.71	27/25/24	-	-	-	-	not meas
		10.1	15.5	CA-1474	3.2	18.2	109.6	76.9	44.2	32.68	39/33/4	-	-	-	-	not meas
		9	40.3	CA-0220	3.0	17.1	103.0	58.1	13.2	44.88	47/16/0	-	-	-	-	not meas
Biochemical	Brix	1_8	16.8	CA-1409	5.6	25.3	9.4	10.7	12.0	-1.31	37/24/27	ns	8.0	-	8.0	60/0/3
composition	Glucose	1_8	16.8	CA-1409	3.5	16.8	3.4	3.8	4.2	-0.40	37/24/27	ns	2.9	-	3.1	60/0/3
		9	16.2	CA-0637	2.9	14.2	3.6	4.6	5.5	-0.94	81/6/1	ns	2.9	3.0	-	45/18/0
	Fructose	9	16.2	CA-0637	3.6	17.1	3.7	4.8	5.9	-1.08	81/6/1	ns	3.1	-	3.1	60/0/3
		1_8	16.8	CA-1409	3.5	16.7	3.5	3.9	4.3	-0.41	37/24/27	ns	3.1	-	3.2	60/0/3
	Malaat	1_8	16.8	CA-1409	3.6	17.2	30.3	38.3	46.4	-8.02	37/24/27	ns	24.2	-	25.2	60/0/3
		1	2.6	CA-0398	3.0	14.5	31.6	40.9	50.3	-9.39	42/37/9	**0.005**	**22.3**	**-**	**64.3**	**60/0/3**
		9	28.5	CA-0558	2.7	13.2	35.3	51.1	67.0	-15.85	77/10/1	ns	24.2	-	25.2	60/0/3
	Citrate	1_8	16.8	CA-1409	5.1	23.2	389.9	458.8	527.8	-68.94	37/24/27	ns	333.9	-	303.8	60/0/3
		12.2	17.7	CA-1548	3.3	15.9	441.9	707.7	973.4	-265.75	85/3/0	ns	331.6	-	348.6	60/0/3
		8.2	0.0	CA-1436	3.1	15.1	444.2	643.3	842.4	-199.09	86/1/1	-	-	-	-	no introgr
	Hexanal	1_8	16.8	CA-1409	12.1	46.8	18.5	19.4	20.2	-0.85	37/24/27	ns	17.6	-	17.4	60/0/3
		4	35.4	CA-1281	3.5	16.6	19.1	20.1	21.2	-1.03	73/14/1	-	-	-	-	no introgr
	3-Hepten-2-one	1_8	16.8	CA-1409	5.9	26.5	16.4	15.9	15.3	0.59	37/24/27	**0.05**	**14.3**	**-**	**13.4**	**60/0/3**
	Linalooloxide	10.1	16.6	CA-0218	11.1	44.1	15.0	17.3	19.6	-2.29	47/41/0	**0.0001**	**13.8**	**-**	**17.1**	**54/0/9**
	p-Menth-1-en-9-al	10.1	16.6	CA-0218	10.2	41.4	16.3	18.5	20.7	-2.24	47/41/0	**0.0001**	**14.9**	**-**	**18.3**	**54/0/9**
		1	20.2	CA-1188	4.1	19.1	16.6	18.0	19.3	-1.35	47/38/3	**0.0005**	**15.0**	**-**	**18.6**	**57/0/6**
	PAE	9	24.3	CA-0507	2.9	14.2	9.5	13.0	16.5	-3.54	83/5/0	-	-	-	-	no comp
	BAE	10.1	17.0	CA-1479	8.1	34.6	10.2	13.3	16.4	-3.08	50/35/3	-	-	-	-	no comp
	Geranylacetone	1_8	16.4	CA-1404	4.1	19.1	15.3	16.1	16.8	-0.76	45/43/0	ns	12.9	-	13.2	60/0/3
		10.1	16.6	CA-0342	0.6[1]	3.0	15.6	15.8	16.1	-0.25	47/36/5	**0.0001**	**12.7**	**-**	**14.5**	**54/0/9**

		BC_2S_1 population										NILs				
Class	**Trait**	**LG**	**cM**	**Marker**	**LOD**	**%EV**	**µA**	**µH**	**µB**	**add.**	**A/H/B**	**Signif[1]**	**mA**	**mH**	**mB**	**A/H/B[2]**
	β-damascenone	10.1	16.6	CA-0218	1.3[1]	6.5	15.8	15.2	14.6	0.58	47/41/0	**0.0001**	**13.1**	**-**	**9.5**	**54/0/9**
	Methoxypyrazine	1_8	16.8	CA-1409	3.4	16.4	17.8	18.4	19.0	-0.58	37/24/27	ns	17.6	-	16.1	60/0/3
		3	33.3	CA-1268	3.2	15.5	18.6	17.8	17.1	0.78	66/15/7	**0.0001**	**17.8**	**17.6**	**16.0**	**45/9/9**
Sensory	Crunchy	6.1	14.7	CA-1338	2.8	20.2	61.7	53.8	45.9	7.89	43/13/0	-	-	-	-	no introgr
attributes	Sticky	1_8	17.8	CA-1169	5.6	36.9	37.1	52.2	67.2	-15.03	31/23/2	ns	18.6	-	19.5	60/0/3
	Tough	11.1	22.0	CA-1513	3.7	26.0	28.9	41.5	54.0	-12.56	37/17/2	ns	19.2	-	21.6	60/0/3
		1_8	16.8	CA-1409	2.8	20.4	27.3	35.1	42.9	-7.80	24/19/13	ns	19.2	-	21.3	60/0/3
	Juicy	1_8	17.8	CA-1169	4.5	31.1	50.8	41.0	31.1	9.84	31/23/2	ns	56.5	-	54.9	60/0/3
	Perfume	3	33.3	CA-1268	2.9	21.1	1.7	3.6	5.4	-1.84	42/10/4	-	-	-	-	split up
	Musty	7.1	0.0	CA-1368	5.0	33.4	0.7	2.6	4.5	-1.87	46/8/2	-	-	-	-	no introgr
	Odor	3	33.3	CA-1268	8.0	38.7	1.4	2.9	4.4	-1.50	56/14/6	-	-	-	-	split up
		9	24.3	CA-0507	4.9	26.0	1.7	4.8	7.9	-3.06	71/5/0	-	-	-	-	split up

The marker with the highest LOD score in the BC_2S_1 population is indicated including its linkage group (LG) and position in cM. Additionally, percentage of explained variance (%EV) at the marker position, estimated (µ, Van Ooijen 2009) or direct means (m), genotype distribution (A/H/B) and estimated additive effect (add.) are given. [1] not significant (ns) [2] no variation for trait (no var), trait not measured (not meas), introgression not present (no introgr), compound not detected (no compound) and split up in multiple attributes (split up). For abbreviations of metabolites see Table 2. Metabolite values represent $\log_2$ values of peak intensities. Significant NIL values are printed bold.

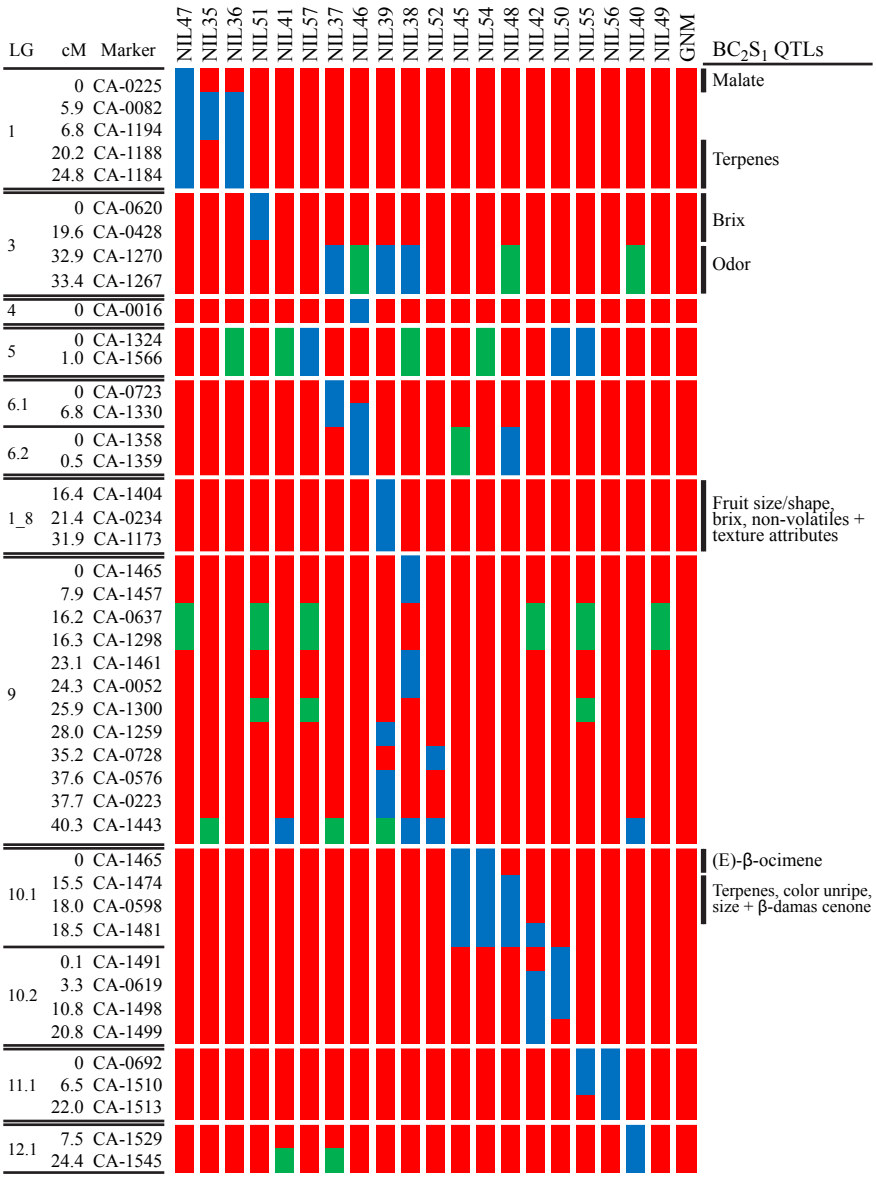


Figure 3. Graphical representation of the NILs and discussed BC_2S_1 QTLs. Only regions with *C. baccatum* fragments are indicated. Introgressions are visualized with their flanking markers in blue (homozygous) or green (heterozygous, i.e. segregating in the NIL).

Table 4 LG3 volatile and flavor QTLs

	BC_2S_1 population						NILs				
Compound/attribute	**LOD**	**%EV**	**µA**	**µH**	**µB**	**Add.**	**Signif.**	**mA**	**mH**	**mB**	**A/H/B**
Aroma	na	na	na	na	na	na	0.0001	51.4	60.8	68.1	45/9/9
Grassiness	na	na	na	na	na	na	0.01	10.6	5.6	5.2	45/9/9
Flowers	na	na	na	na	na	na	0.01	3.8	7.8	8.9	45/9/9
Spices (non-pungent)	na	na	na	na	na	na	0.005	7.0	13.7	15.2	45/9/9
Celery	na	na	na	na	na	na	0.0001	2.2	6.3	8.9	45/9/9
Chives	na	na	na	na	na	na	0.0001	1.4	6.6	10.5	45/9/9
4-Mercapto-4-methyl-2-pentanone	4.3	20.3	11.9	13.8	15.7	-1.89	-	nd	nd	nd	nd
6-Methyl-4-oxo-5-heptenal	16.1	56.9	9.3	13.3	17.2	-3.96	0.0001	10.3	15.0	14.0	45/9/9
unknown_5871	8.1	34.5	13.1	13.7	14.4	-0.64	-	nd	nd	nd	nd
(Z)-Butanoic acid 3-hexenyl ester	3.9	18.3	14.6	13.0	11.3	1.68	0.0001	14.1	13.9	12.0	45/9/9
2-Isobutyl-3-methoxypyrazine	3.2	15.5	18.6	17.8	17.1	0.78	0.0001	17.8	17.6	16.0	45/9/9
(E)-3,6-dihydroxy-2-methyl-1,4-benzoquinone-4-methoxyimine-N-oxide	21.7	67.8	8.7	11.7	14.6	-2.94	-	nd	nd	nd	nd
unknown_8805	17.1	59.2	8.9	11.8	14.7	-2.89	0.0001	8.9	12.5	12.0	45/9/9
unknown_8832	15.2	54.9	8.7	10.7	12.7	-1.98	-	nd	nd	nd	nd
unknown_9961	21.3	67.2	8.7	11.1	13.5	-2.37	-	nd	nd	nd	nd

LG3 refers to the *C. baccatum* introgression of 32.9-33.4 cM from linkage group 3. Percentage of explained variance (%EV), estimated (µ, Van Ooijen 2009) or direct means (m), estimated additive effect (add.) and genotype distribution (A/H/B) are given. Metabolite values represent $\log_2$ values of peak intensities. na= not available, nd= not detected.

LG10.1 and LG1 terpenoid QTLs

An initial analysis of the 137 metabolites detected in the NILs by principal components analysis made clear that a large part of the metabolic variation (46.1%, data not shown) between the genotypes is caused by a group of terpenoids, of which the intensity levels between parents and offspring varied enormously. For the two terpenoids linalooloxide and p-menth-1-en-9-al reported in Table 1, the maximum intensities found in the BC_2S_1 population were up to 39.3 and 19.7 times higher, respectively, than in the PEN45 donor parent. For these terpenoids a common major QTL (LOD>10) on LG10.1 and a p-menth-1-en-9-al specific QTL (LOD 4.1) on LG1 were found (Table 3). Taking a closer look at the NILs having these LG10.1 (NIL45, 48 and 54) or LG1 (NIL36 and 47) introgression revealed a group of at least fifteen terpenoids that are affected (Table 5). In most cases, both introgressions resulted in up-regulation of the compounds with e.g. for α-terpineol having a more than tenfold increase in NILs with the LG10.1 introgression compared to lines without any of these two introgressions. For hotrienol the effect of the introgressions was even larger with a 27.1 fold increase for LG10.1 and a 16.6 fold increase for the LG1 fragment. Some terpenoids were specifically affected by one of the introgressions. For cineole only the LG1 introgression was effective and for (E)-β-ocimene the up-regulation was specific to the telomeric LG10.1 introgression present in NIL45 and NIL54 and absent in NIL48 (Fig. 3). In the BC_2S_1 population the LG1 locus was not significant for ten of the terpenoids, while the introgression in the NILs always resulted in a significant effect. On the other hand the effect of the LG10.1 introgression was always supported by a significant QTL in the BC_2S_1 population (Table 5).

Non-volatile effects in the NILs

As mentioned above, a QTL for malate was found in the BC_2S_1 population of which the effect was confirmed in the NILs. For the other non-volatile compounds glucose, fructose and citrate and the total soluble solids measure Brix, some QTLs were found in the BC_2S_1, but none of these effects could be confirmed in the NILs (Table 3). However, in the NILs themselves we observed some very interesting non-volatile effects, relating to the *C. baccatum* introgression on LG3 of 0-19.6 cM and, to a lesser extent, to the LG10.1 introgression of 15.5-18.0 cM. Brix and the concentration of glucose, fructose and citrate were significantly increased in NILs having these fragments, while in both cases the malate concentration was not significantly affected (Table 6).

The LG3 introgression gave the strongest effects (Table 6), resulting in e.g. a Brix

increase of 1.76 degrees compared to all NILs lacking the fragment and even of 2.47 degrees compared to the recurrent parent GNM. Most interesting, this effect seemed to be unrelated to fruit size, as the fruits of NIL51 had a similar size (8x7 cm; length x width) as GNM (8x8 cm). The LG10.1 introgression, on the other hand, coincided with confirmed length and width QTLs, resulting in fruits of on average 5x5 cm.

Table 5 LG10.1 and LG1 terpenoid QTLs

					BC_2S_1 population					NILs			
Compound	**El. comp**[1]	**LG**[2]	**LOD**		**%EV**	**µA**	**µH**	**µB**	**Add.**	**Signif.**	**mA**	**mB**	**A/B**
α-Terpinene	$C_{10}H_{16}$	10.1	4.3		20.2	13.9	14.8	15.7	-0.88	0.0001	10.7	13.2	54/9
		1	0.0	ns	0.0	14.3	14.3	14.3	0.02	0.001	10.9	13.2	57/6
γ-Terpinene	$C_{10}H_{16}$	10.1	8.3		35.3	12.0	13.7	15.3	-1.62	0.0001	10.4	12.8	54/9
		1	2.3	ns	11.0	12.4	13.4	14.3	-0.94	0.001	10.5	12.9	57/6
Terpinolene	$C_{10}H_{16}$	10.1	10.3		41.7	14.7	16.7	18.7	-2.00	0.0001	12.4	15.3	54/9
		1	2.3	ns	11.1	15.2	16.3	17.3	-1.07	0.001	12.6	15.2	57/6
Limonene	$C_{10}H_{16}$	10.1	10.4		42.0	15.3	16.6	18.0	-1.35	0.0001	13.6	16.4	54/9
		1	0.8	ns	3.9	15.8	16.2	16.6	-0.43	0.005	13.8	16.1	57/6
Myrcene	$C_{10}H_{16}$	10.1	7.3		31.8	12.0	15.1	18.3	-3.18	0.0001	11.7	14.7	54/9
		1	1.7	ns	8.5	12.7	14.2	15.6	-1.45	0.001	11.8	14.8	57/6
(E)-β-Ocimene	$C_{10}H_{16}$	10.1	2.9		14.1	14.3	15.5	16.8	-1.23	0.0005	12.5	15.1	57/6
		3											
Hotrienol	$C_{10}H_{16}O$	10.1	5.9		26.5	14.2	17.2	20.2	-2.98	0.0001	9.7	14.6	54/9
		1	3.0		14.6	14.7	16.6	18.6	-1.96	0.005	10.0	14.0	57/6
p-Menth-1-en-9-al	$C_{10}H_{16}O$	10.1	10.2		41.4	16.3	18.5	20.7	-2.24	0.0001	14.9	18.3	54/9
		1	4.1		19.1	16.6	18.0	19.3	-1.35	0.0005	15.0	18.6	57/6
Geranic-oxide	$C_{10}H_{18}O$	10.1	10.6		42.6	15.8	16.9	18.0	-1.10	0.0005	13.6	14.6	54/9
		1	0.4	ns	2.1	16.2	16.5	16.7	-0.26	0.005	13.7	14.4	57/6
Myrcenol	$C_{10}H_{18}O$	10.1	10.5		42.2	11.5	13.8	16.1	-2.31	0.0001	9.8	13.2	54/9
		1	1.6	ns	8.0	12.1	13.2	14.2	-1.04	0.005	10.0	12.8	57/6

Compound	El. comp[1]	BC_2S_1 population							NILs			
		LG [2]	LOD	%EV	µA	µH	µB	Add.	Signif.	mA	mB	A/B
α-Terpineol	$C_{10}H_{18}O$	10.1	10.7	42.8	17.9	20.2	22.7	-2.44	0.0001	16.4	19.3	54/9
		1	2.9	14.2	18.3	19.5	20.8	-1.24	0.005	16.6	19.2	57/6
Linalool	$C_{10}H_{18}O$	10.1	9.5	39.3	16.9	19.5	22.1	-2.62	0.0001	16.0	18.1	54/9
		1	2.6 ns	12.7	17.4	18.7	20.1	-1.31	0.005	16.2	17.8	57/6
Cineole	$C_{10}H_{18}O$	1	6.8	29.8	12.0	14.7	17.3	-2.62	0.0001	9.2	14.6	57/6
(E)-Linalooloxide	$C_{10}H_{18}O_2$	10.1	11.0	43.9	16.7	19.1	21.5	-2.41	0.0001	15.2	18.7	54/9
		1	2.7	13.0	17.2	18.4	19.5	-1.16	0.001	15.3	18.8	57/6
Linalooloxide	$C_{10}H_{18}O_2$	10.1	11.1	44.1	15.0	17.3	19.6	-2.29	0.0001	13.8	17.1	54/9
		1	2.5 ns	12.2	15.6	16.6	17.7	-1.07	0.001	13.9	17.3	57/6

[1] Elemental composition. [2] LG10.1 and LG1 refer to markers CA-0218 at 16.6 cM and CA-1188 at 20.2 cM, respectively, on the corresponding linkage groups. [3] refers to marker CA-0022 at 6.3 cM on LG10.1. Again percentage of explained variance (%EV), estimated (µ, Van Ooijen 2009) or direct means (m), estimated additive effect (add.) and genotype distribution (A/B) are given. Metabolite values represent $\log_2$ values of peak intensities. ns= not significant.

Table 6 Brix, sugar and acid effects in the NILs

Trait	LG3				LG10.1			
	Signif.	mA	mB	A/B	Signif.	mA	mB	A/B
Brix	0.005	7.94	9.70	60/3	0.0005	7.89	8.81	54/9
Glucose	0.005	2.87	3.71	60/3	0.005	2.87	3.16	54/9
Fructose	0.01	3.07	3.72	60/3	0.0005	3.05	3.37	54/9
Citrate	0.005	323.40	513.43	60/3	0.005	326.98	365.28	54/9
Malate	ns	24.48	19.79	60/3	ns	25.22	18.53	54/9

LG3 and LG10.1 refer to the *C. baccatum* introgressions of 0-19.6 cM and 15.5-18.0 cM, respectively, on the corresponding linkage groups. Direct means (m) and genotype distribution (A/B) are given. ns= not significant.

Discussion

Challenges due to interspecific crossing

A few studies addressed variation in biochemical compounds and agronomical traits in the species *Capsicum baccatum* (e.g. Do Rêgo 2009, Rodriguez-Burruezo 2009, Wahyuni 2013b) and ample suggestions for improvement of genotypes within the species have been made. In this study we have investigated the possibility to use *C. baccatum* for enrichment of the *C. annuum* gene pool. For this purpose the *C. baccatum* var. *pendulum* accession PEN45 was used, as it was shown previously to contain potentially interesting volatile and non-volatile variation (Eggink et al. 2010). Interspecific crossing in combination with embryo rescue was performed to overcome the post-fertilization genetic barriers as described in Yoon et al. (2006). Due to difficulties experienced during interspecific crossing however, we were not able to generate a single bi-parental mapping population of sufficient size, but instead developed the described multi-parent BC_2 and BC_2S_1 populations, which made the genetic analyses more complex.

The total size of the map developed in this study (602.5 cM) is small compared to previously published pepper map sizes by e.g. Wu et al. (2009; 1613 cM) or Barchi et al. (2007; 1857 cM). This is, however, as expected, since our map is based on a BC_2 population with a limited number of effective recombinations, causing clustering of markers. Chromosome five specifically, turned out to be rather underrepresented as the corresponding linkage group consists of only two markers (Fig. 2). Five other markers originating from chromosome five are however present on the map, but positioned on other linkage groups (LG3 (1), LG6.1 (1) and LG9 (3)). An explanation for this could be chromosome five specific translocations, but it seems more likely to be caused by absence of enough polymorphic markers that can be connected in one linkage group. Overall, the available integrated map turned out to be indispensable for orientation and assignment to pepper chromosomes.

Combined QTL detection and validation approach

For the initial experiment 250 BC_2S_1 plants were used originating from 34 BC_2 plants. Depending on the phenotyping protocol plants were measured individually or per plot leading to different numbers of phenotypes per trait (ranging from 56 to 250; Table 1). Consequently, plot genotypes had to be created in cases when (fruits of) multiple plants were combined in the measurements. In this procedure we assumed that the *C. baccatum* effects were dominant, but also took the number of plants in

a plot into account. So e.g. in case a plot contained four plants of which two scored A and two scored H, the plot score would be H (dominance of *C. baccatum* allele), but in case the plot would contain three A scores and one H, the plot score would be A (higher number of A plants). This procedure turned out to work very well as the control traits pungency and red color could be perfectly mapped to the positions at which they were mapped earlier by Thorup et al. (2000; red color) and Blum et al. (2002; pungency). In the initial Interval Mapping analysis for red color, however, four QTLs with limited effect were found (Table 3), of which only the QTL on LG6.2 fell in the expected CCS region. Knowing that a single dominant gene is involved the other three QTLs had to be false positive signals. These artifacts could be explained by accidental co-segregation of the PEN45 allele from markers CA-0323 (LG2.1), CA-0038 (LG12.1) and CA-1209 (LG12.2) with marker CA-0097 from LG6.2 in 5 to 7 BC_2S_1 plants (data not shown), caused by the limited number of BC_2 plants from which the BC_2S_1 plants originate. In addition we showed that due to the multi-parent nature of the BC_2, during map construction, markers linked to fruit color genes were selected against, demonstrating another limitation of our used mapping population. Finally, pre-selection of the BC_2S_1 plants for pungency, color and fertility led to extremely skewed segregations of markers at certain linkage groups, exemplified by the pungency locus at LG2.2 with 247 plants homozygous *C. annuum* versus only 3 plants homozygous *C. baccatum* (Table 3). It should be noted therefore, that these limitations might have resulted in other false positive, or even missed QTLs in the BC_2S_1 population. Nevertheless, there is a vast amount of BC_2S_1 QTLs that were confirmed in the NILs for both physical, biochemical as well as sensory traits, proving the consistency of such QTLs and indicating the strength of our combined approach in multi-parent interspecific mapping populations.

Unexpected flavor variation

As the *C. baccatum* PEN45 accession is pungent it was not included in the sensory evaluations of the expert panel. To base the choice for this accession, as donor parent for backcrossing into *C. annuum,* not only on the volatile and non-volatile variation found (Eggink et al. 2010), fruits of PEN45 were also tested by a group of untrained persons accustomed to eat pungent food. This group described the taste of PEN45 as fruity, not very aromatic and sour. This rather conservative description made the much larger variation in flavors found in the BC_2S_1 experiment, including remarks like ‘tropical flavor’, ‘chives taste’ or ‘similar to papaya’, very unexpected. Especially the flavor effect of the small introgression on LG3 is evident and an enrichment of the taste variation within *C. annuum*. The size of the fragment, 0.5

cM based on the genetic map developed within this study, corresponds to a region of 8.5 cM on the integrated map. This size of an introgression, in combination with the availability of in-fragment markers and absence of (clear) linkage drag, makes it very interesting for breeding. In the BC_2S_1 population an additional odor QTL was found on LG9 at 24.3 cM (Table 3). In the NIL experiment we could however neither confirm, nor completely rule out an effect of this locus alone, as this introgression was only present in NIL38, which also contains the LG3 fragment. Given the population structure, most likely it is a false positive signal due to accidental co-segregation in the BC_2S_1.

We investigated not only the genetics of the observed flavor variation, but also made an attempt to identify the responsible biochemical compounds. The metabolites 6-methyl-4-oxo-5-heptenal and unknown_8805 were confirmed to be specifically up-regulated in the LG3 containing NILs and BC_2S_1 plants, at the same time showing a strong correlation with odor. Given the complex nature of flavor, however, it is unlikely that they are the only responsible compounds. Having therefore a closer look at the six LG3 containing NILs it turned out that the intensity of the two down-regulated compounds, (Z)-butanoic acid 3-hexenyl ester and 2-isobutyl-3-methoxypyrazine, is only decreased in the homozygous NILs (NIL37, 38 and 39) and not in the heterozygous NILs (NIL40, 46 and 48; Table 4). It could be true therefore that these compounds interact with, or even partially mask the effect of, the up-regulated compounds resulting in a more intense aroma in the homozygous NILs (68.11; Table 4) and a slightly suppressed aroma intensity (60.81) in the heterozygous lines, while the concentrations of 6-methyl-4-oxo-5-heptenal and unknown_8805 are even slightly higher in the heterozygous versus the homozygous NILs. Such a suppression effect might also be expected from 2-isobutyl-3-methoxypyrazine, as it is commonly described in sniffing port analyses as characteristic (green) bell pepper aroma (Luning et al. 1994a, Van Ruth et al. 1995, Rodriguez-Burruezo et al. 2010), while the LG3 NILs are especially described as having non-typical pepper aroma, with use of the attributes flowers, spices, celery and chives. An effect of 2-isobutyl-3-methoxypyrazine alone on aroma was not found, as NILs with a significantly decreased (NIL47 and NIL56) or increased (NIL45 and 54) intensity were unaffected for aroma intensity.

A follow-up experiment with sub-NILs having either LG3 or LG9 and the combination of both would serve validation of the genetics and at the same time further elucidation of the biochemical relations.

Terpenoids and their potential application

In both mapping populations a large variation in terpenoid levels was found, with for some terpenes a maximum concentration, which was almost 40 fold higher than detected in the most extreme parent. The described PEN45 LG10.1 and LG1 introgressions had major effects on the terpenoid content of the mature fruits, affecting at least fifteen different terpenoids (Table 5). These compounds, derived from precursors IPP and DMAPP, belong to a large class of terpenoid metabolites that serve multiple roles in plants and which are sub-divided in the main groups monoterpenes (C10), sesquiterpenes (C15), and diterpenes (C20). In our study, the QTLs on LG10.1 and LG1 affected the accumulation of monoterpenes only, whereas sesquiterpenes and diterpenes were unaffected by these two introgressions. In the BC_2S_1 population, 12 sesquiterpenes were detected however, of which 4 compounds (γ-himachalene, α-cuprenene, δ-amorphene and α-muurolene) gave significant QTLs on LG1_8 and LG8 (data not shown). Unfortunately, neither of these two regions were represented in the NILs and could therefore not be validated.

The backbones of mono-, sesqui- and diterpenes are synthesized by enzymes that belong to the structurally related terpene synthase (TPS) family (Bohlmann et al. 1998). The genome of cultivated tomato (*Solanum lycopersicum*) contains 44 of such TPS genes, including 29 that are functional or potentially functional. Many of the tomato TPS genes are found in clusters on chromosomes 1, 2, 6, 8, and 10 (Falara et al. 2011). Chromosomes 1 and 10 of pepper are highly syntenic to those of tomato (e.g. Wu et al. 2009), which make the tomato TPS genes in the chromosome 1 and 10 clusters interesting candidates for the monoterpenoid concentration increase caused by our PEN45 introgressions. Furthermore, this matches with the observation that the majority of the TPS genes in the tomato chromosome 1 and 10 clusters encode monoterpene synthases (Falara et al. 2011).

Although, some monoterpenes are well-known flavor compounds (e.g. linalool and β-ocimene; Luning et al. 1994a) in our study there was no co-localization of terpenes and any taste attribute. Still, the increase in monoterpenoid concentration can be relevant for pepper, since terpenoids have been shown to play a role in the interactions of plants with their environment. More specifically, they can be active as direct defense compounds, containing antifungal or antibacterial properties (reviewed by Kalemba and Kunicka 2003). In addition, volatile terpenoids can function as indirect defense compounds by attracting predators or parasitoids of the attacking insect (reviewed by Walling 2000). For some of the elevated monoterpenes in this study (Table 5) specific relations with relevant pepper pathogens have already been described. E.g. in cucumber (*Cucumis sativus*), although in vegetative organs

instead of fruits, a positive correlation was found between the attraction of predatory mites (*Phytoseiulus persimilis*) and the amount of emitted (E)-β-ocimene, after infestation of the plants with herbivorous spider mites (*Tetranychus urticae*; Kappers et al. 2011). In addition, antimicrobial properties related to monoterpenes have been reported in several studies of essential oils. Perez-Sanchez et al. (2007) e.g. reported a clear growth inhibition of the pathogenic fungi *Colletotrichum acutatum* and *Fusarium oxysporum* (causing anthracnose and internal fruit rot, respectively, in pepper), which showed the highest correlation with the concentration of the monoterpene α-terpinene, extracted from the oil of *Thymus zygis*. These examples indicate that it will be very interesting to study the behavior of the LG10.1 and LG1 containing NILs in relation to pathogen infestation.

Like for the LG3 flavor introgression, the size of the LG10.1 and LG1 fragments is limited, i.e. 2.5 and 4.6 cM, respectively, corresponding to 17.2 and 10.3 cM on the integrated map. It should be noted however, that the LG10.1 introgression also contains QTLs affecting length, width and unripe fruit color, resulting in fruits which are smaller (on average 5x5 cm) than the recurrent parent GNM (8x8 cm) and having a pale green color. In addition, NILs with the LG10.1 introgression show a decreased concentration of β-damascenone, a product of β-carotene degradation, while geranylacetone, which is produced from open chain carotenoids, precursors of β-carotene, shows up-regulation at that locus (Table 3). Brand et al. (2012) also described a QTL for unripe fruit color at the top of chromosome 10 in a cross between a dark-green *C. annuum* and a pale green *C. chinense* accession. They showed that the unripe fruit color intensity reflects the content of chlorophyll and other metabolites associated with the chloroplast at immature fruit stage, and is likely due to the increased chloroplast number and chloroplast compartment size observed in dark-green genotypes, as found in high pigment tomato accessions (Cookson et al. 2003). Although dark-green genotypes also had more and larger chromoplasts at ripe stage, levels of the major carotenoids accumulating at ripe stage seemed not to be affected. On the one hand, we cannot rule out the possibility that chloroplast abundance and morphology rather than variation in biosynthetic pathway genes may be the cause of the QTLs for terpenoids and carotenoid-derived volatiles in our LG10.1 NILs, since these compounds are synthesized in plastids (Nagegowda 2010). On the other hand, as mentioned, it is known that the upper part of tomato chromosome 10 contains a cluster of monoterpene synthase genes (Falara et al. 2011), which would also be likely candidates for the observed variation in terpenoids. The availability of a pepper genome sequence will certainly be of great help to address which candidate genes could be involved in controlling the traits on LG10.1.

In contrast to the LG10.1 NILs, fruits of NIL36 and 47, containing the LG1

introgression affecting terpenoid content, have a similar size and fruit color as the recurrent parent, making them amenable for direct use in breeding.

Fruit size independent Brix increase

For the non-volatile compounds glucose, fructose, malate and citrate and the total soluble solids measure Brix a common QTL with positive effect from the *C. baccatum* allele was found on LG1_8, which co-localized with a QTL affecting the texture related attributes toughness, stickiness of the skin and juiciness (Table 3). As these QTLs coincided as well with the main fruit size QTL, which led to severely reduced fruit width and conical shape in the BC_2S_1 population, it seems plausible that the increase in the concentrations of the non-volatiles resulted directly from smaller fruit size, rather than an absolute increase in amount. A negative relation, that is well-known in e.g. tomato, where small size cherry tomatoes have higher sugar concentrations than large-fruited beef tomatoes. The same could be argued for the texture related attributes, as smaller conical fruits had generally more rigid fruit walls leading to less juicy and tougher fruits. Both seem to be in line as well, with the observation that in the NILs, where the LG1_8 QTL had only a minor effect on fruit width, the non-volatiles and texture attributes were not significantly affected anymore (Table 3).

In the NILs, however, we did find some interesting regions with an effect on non-volatile concentrations, as the *C. baccatum* introgression at the top of LG3 in NIL51 and, to a lesser extent, the LG10.1 fragment of 15.5-18.0 cM in NIL45, 48 and 54 resulted in an enormous increase in Brix and underlying soluble solids, glucose, fructose and citrate; compounds which are known to play an essential role in flavor. In tomato e.g., taste intensity is mainly attributed to reducing sugars and organic acids (Stevens et al. 1977, Krumbein and Auerswald 1998, Bucheli et al. 1999) and overall liking, as determined by consumer panels, is found to be strongly correlated with Brix levels in both pepper and tomato (Verkerke and Kersten 2000). Often, however, a negative relationship between Brix and fruit size and/or yield is found (Grandillo et al. 1999, Georgelis et al. 2004). Also in the LG10.1 NILs such a negative effect on fruit size was found, making them less interesting for breeding. It is remarkable therefore that the LG3 fragment with the largest effect on total soluble solids content did not show this negative relation, as NIL51 fruits were similar in size as those from the recurrent parent. A next step will be to perform a solid yield experiment to study the effect on total harvest.

It is not completely clear why in the BC_2S_1 population we did not find any Brix QTL on LG3, while the effect in the NIL population is evident. A plausible

explanation could however be, that in the BC_2S_1 population still multiple genetic factors influencing Brix levels, like sugar and acid synthesis, transport and conversion, but also factors like fruit size and fruit set, segregated simultaneously, which, in combination with aforementioned limitations of our mapping population, made mapping of individual effects impossible. It anyhow proves, that next to the validation possibilities, the NILs were indispensable for dissecting a complex trait like Brix.

Conclusion

This study demonstrates that *Capsicum baccatum* is a valuable source for enrichment of the *C. annuum* genetic pool. In several cases unexpected traits were introgressed in *C. annuum*, as shown by the wide flavor variation, transgressive terpenoid levels and the fruit size unrelated Brix effect. The combination of developed populations allowed mapping of simple morphological traits, but also genetic dissection of quantitative traits with complex inheritance patterns. The current *C. annuum* genome sequence knowledge in combination with candidate genes from syntenic crops, like tomato, will allow further elucidation of our studied traits. Furthermore, the availability of NILs containing a limited amount of introgressions of restricted size, make flavor enhancement and/or potential (in)direct defense applications in commercial breeding programs directly possible.

Acknowledgements

The authors kindly acknowledge Harry Jonker and Yvonne Birnbaum at Plant Research International for performing the SPME-GC-MS analyses. In addition we thank Suzanne de Wit, Femke Willeboordse, Laure Pohu, Sander Bos, Tineke Benning and Paula de Grauw for performing a massive job on sample preparation and non-volatile measurements. Finally we are grateful to all people at Rijk Zwaan who took care of perfect greenhouse management.

Chapter 6

General discussion

In the breeding process of pepper, production (yield) and quality, such as shelf life, firmness and disease resistance, are traits of main interest. Consumer acceptance of vegetables is, however, highly dependent on appearance and flavor (Rocha et al. 2013). More than a decade ago, improvement and diversification of pepper flavor have been identified already as chances for added value creation and consumer satisfaction (Verkerke 2000). Flavor of fruits and vegetables, as perceived during consumption, is however a complex trait which has been defined as the overall sensation provided by the interaction of taste, odor, mouth feel, sight and sound (Luning et al. 1994b). Especially the interplay among all these parameters in combination with different consumer preferences makes flavor such a difficult property to quantify in an objective way, which is a prerequisite for successful flavor breeding.

In order to pave the way to a better understanding, prediction and improvement of pepper flavor, at the beginning of this project we formulated three main objectives (General introduction). We planned to characterize and identify volatile and non-volatile compounds of *Capsicum* in relation to flavor and yield and study the genetics of these compounds. In addition it was our objective to introduce identified (non-)volatile compounds in pre-breeding lines for improvement of overall flavor with use of marker assisted breeding. In this chapter obtained results, insights and consequences for pepper flavor (breeding) are discussed in relation to the afore mentioned objectives.

Plant material

This PhD project started with the composition of a diverse panel of genotypes that represented, (i) roughly the flavor variation in the commercial *Capsicum annuum* breeding program of Rijk Zwaan, (ii) parents of available mapping populations and (iii) some genotypes that were expected to have extraordinary flavors. As the outcome of a project highly depends on the input, some rationale behind inclusion of genotypes in our starting panel is given here.

The complete panel consisted of 35 genotypes, of which 20 genotypes belonged to the first class, representing the flavor variation in elite breeding lines and commercial hybrids (indicated as 'elite' in Table 1, Chapter 2). Fruits ranged from 5-22 cm in length and 4-8.5 cm in width, within the fruit types blocky, dulce italiano, dolma and kapya. The majority of the genotypes were red, as this is the predominant color in cultivated material; yellow and orange genotypes were less represented. At the start of the project four relevant mapping populations, of the types RIL (recombinant inbred line) and BIL (backcross inbred lines), were available from which the parents (Table 1) were included in the genotype panel. Worth mentioning with respect to the

C. baccatum pendulum PEN45 population, is line J, described in Chapter 2, 3 and 4, which corresponds to line GNM, that is the recurrent parent of the BIL population described in Chapter 5.

Table 1. Mapping populations available at the start of the project

Type	Donor parent	Recurrent parent	Generation
RIL	C. frutescens BG 2814-6	C. annuum NuMex RNaky	F_7
BIL	C. baccatum pendulum PEN45	Several C. annuum breeding lines[1]	BC_2S_1
BIL	C. baccatum pendulum PEN79	C. annuum Vania[2]	BC_4S_2
BIL	C. annuum CM334	C. annuum Maor	BC_3S_2

[1] For description of the population see Chapter 5. [2] *C. annuum* Turrialba was used as bridge in the first cross with PEN79.

The third class of genotypes, with expected extraordinary flavors, contained the *C. annuum* genotypes Piquillo, Buran, PBC1405, PI 543188 and Perennial and the *C. chinense* genotypes Antillais Caribbean and Chinense-WA. The cultivated variety Piquillo was chosen as it is famous in the Mediterranean region for its full taste and rich aroma. Cultivar Buran was reported to be a very sweet lamuyo type (http://www.seedsavers.org) and Perennial was chosen because it is a very wild *C. annuum* accession, potentially harboring flavors of interest. PBC1405 and PI 543188 were included as they were reported to be non-pungent accessions of non-*C. annuum* origin, which is rather unique as most accessions from wild *Capsicum* species are pungent. According to the AVRDC gene bank (http://www.avrdc.org) PBC1405 had to be a non-pungent *C. baccatum* accession and according to the genebank of the USDA (http://www.ars-grin.gov) PI 543188 was supposed to be a non-pungent *C. chinense* accession. In our experiment however, both accessions turned out to belong to *C. annuum*. Finally, the species *C. chinense*, which is specifically mentioned in literature to harbor unique combinations of flavors and especially aromas (Bosland and Votava 1999), was represented by the very aromatic cultivars Antillais Caribbean (www.technisem.com) and Chinense-WA (obtained from a local breeder in West-Africa). Chinense-WA segregated for yellow and red fruits. Biochemical measurements and sensory evaluations of this cultivar were therefore performed on samples of the separate fruit colors.

The panel turned out to be a diverse set in terms of agronomical evaluations (type, size, color, pungency and yield), but especially in terms of biochemical and sensory variation (Chapters 2-4). From the mapping populations (Table 1) the two *C. baccatum pendulum* BILs were identified as interesting candidates for further study because of the elevated acid concentrations and aberrant volatile profiles of PEN45

and PEN79. Compared to e.g. Mazurka the citrate concentration of the *C. baccatum pendulum* accessions was 2.5-3 times higher and the malate concentrations were even up to 12 times higher (Table 2, Chapter 2). The biochemical profiling with use of SPME-GC-MS allowed visualization of between- and within-species volatile compound variation. Principal components analysis (PCA) on the intensity patterns of 391 putative volatile compounds revealed individual grouping of *C. chinense, C. baccatum* var. *pendulum* and *C. annuum*, indicating potentially interesting volatile variation present in, not only, the *C. baccatum pendulum* group, but also in the species *C. chinense* (Fig. 2, Chapter 2). A large group of saturated and unsaturated esters were mainly responsible for the individual grouping of the *C. chinense* accessions (Chapter 2). An interesting finding, as this type of compounds obtained fruity, sweet and exotic odor descriptions by a trained sniffing port panel (Rodriguez-Burruezo et al. 2010). Populations with the *C. chinense* accessions Antillais Caribbean and Chinense-WA were, unfortunately, not available. The *C. annuum* CM334 x Maor BIL was not chosen for further study, as both parents were not extremely deviating from the elite material for both their biochemical profiles and sensory evaluations (see next paragraph). The *C. frutescens* BG2814-6 x *C. annuum* Numex RNaky RIL was ignored as well, since the flavor of BG2814-6 was described as very bitter (see next paragraph).

Sensory aspects

From the complete set of 35 genotypes (Chapter 2), 24 genotypes were non-pungent (Chapter 3 and 4). The number of non-pungent genotypes was chosen in such a way that it was possible for the descriptive sensory panel to evaluate them in two subsequent days, with the three repetitions of the genotypes split over multiple sessions. To further increase the number of genotypes was not possible, because this would have resulted in a longer period of sensory trials with the potential risk of quality decay of the fruits during storage. A similar limitation seems to be encountered by others as well, as sensory studies on (processed) fruit vegetables are often performed on a limited amount of genotypes; e.g. Sinesio et al. 2010 (16 tomato varieties), Rodriguez-Burruezo et al. 2010 (16 *Capsicum* accessions) or Vallverdú-Queralt et al. 2013 (12 tomato juices). Yet in our case, to increase the number of data points for the statistical analyses, we decided to evaluate the same genotypes multiple times during the growing season, which resulted in a data set with evaluations on 24 genotypes in 3 repetitions over 3 harvests (minus some missing values). The three repetitions of the genotypes within the trial allowed to calculate percentages variance explained by genotype. So, although we did not run markers on our panel, still we were able to

make an estimation of the heritability of the taste attributes (Chapter 3 and 4), which, to our knowledge, is done for the first time in pepper.

We analyzed the sensory data using a linear mixed model REML (residual maximum likelihood) analysis. Mean values were calculated per genotype per replicate after a correction for session and panelist effects and removal of strong outliers (if the absolute value of a standardized residual was larger than three residual standard deviations). From this analysis it became clear that, even after removal of the strong outliers, a large part of the observed variation in flavor scores could be attributed to the taste panel itself, other than to the material, although this varied for the individual taste attributes. In Chapter 3 we reported that the estimated percentage of variance attributed to genotype, ranged from 0% to 47.1% for musty, carrot, petrochemical and green bean taste versus juiciness, respectively. Consequently, the percentage of variance attributed to session and panelist ranged from 100% to 48.5% (Table 3, Chapter 3). It can be concluded therefore that the sensory evaluations form a large source of variation within our experiments and that not all taste attributes are equally suitable for consistent scoring by a taste panel. This makes it both very useful to replace the sensory evaluations by indirect predictors of taste, but also very difficult to first identify reliable predictors.

Although the 11 pungent genotypes could not be evaluated by the descriptive sensory panel, still we tried to acquire as much sensory data as possible, by addressing the pungent genotypes to a panel of mainly Asian people accustomed to eat pungent food. This newly trained panel described the flavor with special emphasis on odor (Table 2), but were not able to quantitatively score the separate attributes like the descriptive panel did. This analysis confirmed the selection of the *C. baccatum pendulum* and *C. chinense* accessions to be interesting candidates for further study; PEN45, specifically, was described as having a fresh, sour odor and a fruity, slightly aromatic and sour taste, while Antillais Caribbean and Chinense-WA yellow were described as being aromatic with fruity, sweet and sour flavor notes.

For the description of the 24 non-pungent genotypes in Chapters 3 and 4 the panelists used fourteen attributes to describe the flavor sensation in the mouth/throat, which were the texture attributes crunchiness, stickiness of the skin, toughness and juiciness, the basic taste attributes sweetness and sourness and the retronasal flavor attributes aroma intensity, grassiness, green bean, carrot, fruity/apple, perfume, petrochemical and musty. However, during the sensory evaluation of fruits from the PEN45 derived BC_2S_1 material (Chapter 5), it became clear that the vocabulary (i.e. predefined attributes) of the trained panel was not sufficient to cover all the flavor variation. This resulted in remarks on the evaluation sheets like ‘presence of tropical fruit flavor’ or ‘similar to papaya taste’, which indicates that we have developed

material with truly new flavors that, so far, were not present in commercially available genotypes. To fully exploit and characterize this new flavor variation, the sensory panel was re-trained, preceding the experiment with PEN45 near-isogenic lines (NILs), with fruits from preselected NILs having more extreme flavors, which resulted in an expansion of the panel's vocabulary with the attributes flowers, spices (non-pungent), celery, chives and bitter. In the subsequent NIL experiment significant genotype differences were actually found for all of these. This indicates that development of genetic material with new flavor variation requires extra training of sensory panels in order to remain up to date.

Table 2. Flavor description of pungent genotypes

Genotype	Odor	Pungency	Taste description
NuMex RNaky	Fresh, pepper	Mild	Sweet, with bite, very juicy, bit fresh and fruity
Line P	Fresh, pepper	Mild	Fresh, sweet, not aromatic
PI 543188	Spicy, pepper	Mild	Fleshy, bit bitter, not sweet, dry, pepper taste
CM334	Green and spicy	Mild	Juicy, not very sweet, no special taste
PEN45	Fresh, sharp, sour	Hot	Fruity, bit aromatic and sour
PEN79	Pungent	Hot	Dry, bit sweet, green and fresh aroma, sour
Antillais Caribbean	Sour, flowers, aromatic	Extremely hot	Aromatic, sweet and sour, fresh and juicy
Chinense-WA red	Sour and pungent	Very hot	Crunchy, bitter, grassy, dry
Chinense-WA yellow	Sour and pungent	Extremely hot	Sweet, fruity aroma, fleshy, bit bitter
Perennial	Sharp	Extremely hot	Thin fruit wall, very bitter, not aromatic
BG2814-6	Hot pepper	Hot	Very thin fruit wall, very bitter
Turrialba	Fresh	Hot	Bitter, not aromatic

Random Forest

The metabolomics approach that we used yielded a data set with more variables (metabolites) than samples (genotypes). In statistics this is known as the classical p>n problem. To tackle this problem we compared several multivariate linear regression approaches with use of OmicsFusion (http://www.plantbreeding.wur.nl/omicsFusion), which is a tool that uses different regression methods to analyze typical ~omics data sets with large numbers of variables and smaller numbers of samples. OmicsFusion applies univariate regression and the regularized multiple regression methods ridge regression (RR), Lasso, elastic net (EN), principal components regression (PCR), partial least squares regression (PLS), sparse partial least squares regression (SPLS) and Random Forest regression (RF) to analyze data sets. In the

methods PCR, PLS, RR and RF all variables are used, while Lasso, EN and SPLS make use of variable selection, resulting in predictions with a smaller number of variables than present in the total set. The generated output contained the mean square error of prediction, goodness-of-fit, variable selection (for those methods that perform variable selection) and ranking of the variables (per method and over the methods). Often the same compounds showed importance in different regression methods, but the number of selected variables, in approaches making use of variable selection, varied a lot. We decided to choose an approach without variable selection, in order to judge the results of all metabolites. Finally, from these approaches we selected Random Forest regression, because good experience was obtained with it at Wageningen UR Plant Breeding in (e.g.) a potato project with complex ~omics data (Acharjee et al. 2011).

Random Forest is an ensemble learning method for classification and regression that operates by constructing a multitude of decision trees at training time and outputting the class that appears most often of the classes yielded by individual trees (Breiman 2001). A Random Forest model is typically made up of hundreds of decision trees. Each decision tree is built from a bootstrap sample of the original data set, bootstrap having the same size as the original sample (n identical; sampling with replacement). That is, some of the original samples will be included more than once in a particular bootstrap sample, whereas others will not appear at all. Generally, about two thirds of the samples will be included in a bootstrap sample and one third will be left out (called the out-of-bag samples). The performance of the models is expressed by the prediction $R2$, which is calculated from the out-of-bag samples. This $R2$ value therefore is not a goodness-of-fit of the data at hand but an estimate of predictive accuracy on independent (left-out) samples. This makes the approach for our type of (prediction) work more powerful, compared to e.g. simple linear regression techniques in which the predictive value of the $R2$ values are still not known. In addition, importance of the individual variables could be determined from the increase in mean square error (MSE) after permutation, which allowed us to identify the metabolites that were most essential in the predictions. A limitation of the used approach, however, is that in cases when there are multiple (often correlated) compounds with a similar contribution to an attribute, not all of them are significant. In our case we determined significance of the prediction $R2$ and variable importance by another permutation test, in which the sensory attributes were individually permuted over the genotypes while retaining the original metabolic values of the genotypes. In this thesis we have shown that Random Forest regression can be successfully used for linking sensory attributes to metabolite data (discussed in next paragraph) and for prediction of attributes within (Chapter 3) and between harvests (Chapter 4).

Metabolites and flavor

In accordance to the genetic diversity of our studied panel, the 24 non-pungent genotypes displayed a high degree of organoleptic and metabolic variation (Chapter 3 and 4). With use of GC-MS in combination with multivariate mass spectral reconstruction we could distinguish 254 different volatile compounds in these genotypes over the three harvests. This number of volatiles is very comparable to the number of volatiles found in twelve *C. annuum* genotypes by Rodriguez-Burruezo and colleagues (2010). For the complete set of 35 genotypes, however, a much larger number of volatiles (391) was found, indicating that the pungent genotypes on the one hand form a good complement to the non-pungent types in our study and on the other hand conceal a metabolic wealth of at least 137 metabolites, with potentially interesting flavor characteristics. In this thesis, we started to explore this variation by the characterization of the *C. baccatum pendulum* derived populations (Chapter 5 and General discussion paragraph *Flavor from C. baccatum*).

Based on the non-pungent genotypes, we found highly correlated clusters of volatiles and non-volatiles, which could be related to metabolic pathways and common biochemical precursors (Chapter 3). Contrasts between genotypes were caused by both qualitative and quantitative differences in these metabolic clusters, with the phenolic derivatives, higher alkanes, sesquiterpenes and lipid derived volatiles forming the major determinants. It seems likely therefore, that changes in genes or expression of genes in such pathways would alter the composition of complete clusters of volatiles, thereby affecting individual attributes or even overall flavor, as was demonstrated in tomato and watermelon (e.g. Lewinsohn et al. 2005, Tieman et al. 2006). In line with this, we speculated in this thesis, based on the knowledge obtained from the *C. baccatum pendulum* PEN45 derived material, that one or more terpene synthase (TPS) genes are responsible for the huge changes in monoterpenoid concentrations observed in specific PEN45 NILs (Chapter 5).

To relate the sensory attributes to the metabolite data and to determine the importance of the individual compounds, as discussed, we used Random Forest regression on the individual harvests and on the three harvests together. Several predictors for the attributes aroma, fruity/apple, sourness and sweetness were found in common between harvests (summarized in Table 3), which we propose as key-metabolites involved in flavor determination of sweet pepper (Chapter 4). This list contains compounds with known relations to attributes, like sweetness and sugars, but also several compounds with new relations. In this respect we have demonstrated for the first time, that the metabolites p-menth-1-en-9-al, (E)-β-ocimene, (Z)-2-penten-1-ol, and 1-methyl-1,4-cyclohexadiene are related to fruity/apple taste

and/or sweetness of pepper. These results are supported by the flavor description of these compounds in combination with the direction of their correlation to the respective attribute (Table 3). The only exception to this is (E)-β-ocimene, which showed subsequently a positive and negative correlation with sweetness in the first and third harvest. Striking for this compound is that it received also contradicting flavor descriptions in earlier performed studies, i.e. fish, rotten (Van Ruth et al. 1994) versus sweet, herbal (http://www.thegoodscentscompany.com). As physical properties of the fruits other than dry matter content were not measured, predictors for the texture related attributes juiciness, toughness, crunchiness and stickiness are not reported, though several well correlated metabolites were found. The intensity differences of these metabolites are, however not causal for, but more likely a result of the texture differences. For the less contrasting attributes grassiness, green bean, carrot, perfume, petrochemical and musty flavor, the only interesting finding was 2-pentyl-furan with a significant contribution to the prediction of green bean flavor. The flavor of this compound was previously also described as green bean/butter like (http://www.flavornet.org/d_odors.html).

Table 3. Sweet pepper key-metabolites with predictive value in multiple harvests

Compound	Attribute[1]	Corr[2]	Flavor description	Reference
(E)-2-Hexen-1-ol	a/f/s	+	Almond, fruit, spicy	Luning et al. (1994a)
Neopentane	a/f/s	+	-	
Fructose	a/f/s	+	Sweet	
p-Menth-1-en-9-al	f/s	+	Spicy, herbal	Good scents company[5]
Glucose	f/s	+	Sweet	
3-Hepten-2-one	f	-	Mushrooms	Van Ruth et al. (1994)
(E)-β-Ocimene	s	+/-[3]	Sweet, herbal	Good scents company
			or: fish, rotten	Van Ruth et al. (1994)
unknown $C_{15}H_{24}$	s	-	-	
(Z)-2-Penten-1-ol	s	-	Rubber, plastic, green	Flavornet[6]
1-methyl-1,4-Cyclohexadiene	s	+	Fruity	Good scents company
unknown $C_{13}H_{18}$	s	-	-	
unknown $C_6H_8O_2$	s/so	-/+[4]	-	

[1] Involved attributes aroma (a), fruity/apple (f), sweetness (s) and sourness (so). [2] Direction of Pearson's correlation coefficient. [3] (+)correlation in first harvest, (-)correlation in third harvest. [4] (-)correlation with sweetness, (+)correlation with sourness. [5] http://www.thegoodscentscompany.com , [6] http://www.flavornet.org/d_odors.html.

Flavor is a very complex trait, we realize therefore that the lists of compounds with significant contribution in attribute prediction are neither exhaustive nor unique. Neopentane, for example, was found to contribute to both aroma and fruity/ apple taste as well as sweetness, however no flavor description for this compound was found. The most likely explanation for neopentane being correlated is, that in the genotypes it has a very similar intensity profile as (E)-2-hexen-1-ol, which seems truly predictive based on its almond, fruit, spicy odor description (Luning et al. 1994a). On the other hand, we would like to emphasize that it is likely that, due to the Random Forest regression approach we used, in some cases there are more (often correlated) compounds contributing to the attributes, but only the best one or two predictors from such compound cluster are listed. For example, in the case of aroma, in addition to (E)-2-hexen-1-ol also 2-hexenal, originating from the same lipid derivative cluster and having fruity, almond odor notes (Luning et al. 1994a), is strongly correlated to aroma, but not listed in one of the Chapters. Also in our analyses we did not find an effect on flavor attributes of the well-known compound 2-isobutyl-3-methoxypyrazine, which is commonly described in sniffing port analyses as characteristic (green) bell pepper aroma (Luning et al. 1994a, Van Ruth et al. 1995, Rodriguez-Burruezo et al. 2010), while a lot of variation for this compound was found between the genotypes (max/min > 165 within the first harvest). This seems to indicate that sniffing port analyses have different sensitivity compared to the taste evaluations we performed.

For sourness the only compound with a consistent significant contribution was an unknown $C_6H_8O_2$ compound (Table 3). Luning et al. (1994b) reported however, that concentrations of citric and ascorbic acids showed close relationships with the attribute sourness, whereas malic acid was negatively correlated. Taking a closer look at the organic acids in our study (citric, malic and ascorbic acid) showed that citric acid displayed the best relation with sourness, although the correlation (0.34) was not significant. We concluded therefore in Chapter 3, that organic acids do not play a role of importance in determination of sourness in pepper. It also seems reasonable that, this conclusion, which is based on 24 genotypes, can be more generally applied to pepper, since the findings from Luning et al. (199b) were based on only 2 genotypes, the blocky pepper cultivars Mazurka and Evident. In tomato, on the other hand, a correlation of 0.76 was found between titratable acids (mainly citric and malic acid) and sourness in a study with 12 tomato genotypes (Tandon et al. 2003), while the variation and concentration of organic acids in that study was similar to that found in our pepper collection. We postulated therefore, the hypothesis that in pepper the effect of sourness related metabolites is masked by other volatile and non-volatile compounds or texture differences (Chapter 3). Subsequently in Chapter 4

we described, for the first time, a clear sweetness-sourness interaction in pepper and demonstrated that the masking effect of fructose and other sugars explained why we did not find organic acids contributing to the prediction of sourness. The difference between tomato and pepper in this respect, might be explained by their different fruit structure. Tomato fruits are composed of distinct edible tissue types including pericarp and locular gel, that are independently perceived in sensory experiments and which can considerably differ in chemical composition (e.g. Moretti et al. 1998), while the only edible part of pepper fruits is the pericarp.

Flavor prediction

The variation in flavor of the non-pungent genotypes could be reduced into two major sensory contrasts, which were a texture related contrast and the basic sweet-sour contrast. The structure of the PCA plots resulting from the analysis with one harvest (Chapter 3) and the analysis with the combined three harvests (Chapter 4) remained almost identical, indicating the stability of these contrasts. In tomato a similar sweet/fruity versus sour/watery contrast was found in both a study with 16 cultivars (Sinesio et al. 2010) and a study with 94 varieties (Hageman et al. 2010), with a texture related contrast, opposing firmness and mealiness, in the second principal component. Although we found similar sensory contrasts as in tomato, in pepper the texture contrast explained the largest part of variation (~50%), whereas in tomato the sweet-sour contrast was most discriminative (~45% explained variation).

When we tried to predict these major sensory attributes within and between harvests, we noticed that the Random Forest predictions of the texture related attributes (juiciness, toughness, crunchiness and stickiness of the skin) and sweetness were very good. The predictions of the attributes aroma intensity, sourness and fruity/apple were somewhat lower and more variable between harvests, especially in the second harvest (Chapter 4). In general, it can be concluded that prediction of attributes with higher heritabilities works better and is more consistent over harvests. At that stage, we also noticed a major limitation in the measurements we performed, as physical properties of the fruits, like firmness and flexibility of the fruit flesh or skin, were not measured, whereas the texture contrast was most discriminative in overall flavor variation and heritabilities of texture attributes were good (moderate to high). The only physical trait we actually determined was dry matter content of the fruits. This single character correlated already well with the texture attributes stickiness, toughness and juiciness (Chapter 3). In retrospect, it would have been a very useful complementation of our study therefore to perform texture analyses on the fruits. In addition, consumer liking data of our genotypes would have made

it possible to predict overall flavor liking, supplementary to the prediction of the individual attributes. For consumer liking studies, however, large panels (n>100) are required to obtain representative outcomes. This made it impossible to include such studies in our experiment, as the number of plants per genotype that we grew (three repetitions of five plants) was not sufficient to simultaneously harvest enough fruits for such a panel. On the other hand, based on the results obtained in this thesis, we can speculate already that sugar and texture measurements, perhaps complemented with some of the key-metabolites, will be essential elements of a model predicting overall liking of pepper. Currently we are working on the development of such a general pepper taste model within a Dutch research consortium (TTI-GG; http://www.groenegenetica.nl).

Flavor from *C. baccatum*

As mentioned before, we selected the two *C. baccatum pendulum* BIL populations for further study because of the elevated acid concentrations and aberrant volatile profiles in combination with a positive flavor description of especially PEN45. Although at the start of this project we selected both populations, in this thesis we only report the results of the PEN45 derived material (Chapter 5), whereas both populations were actually grown, evaluated for flavor and biochemically characterized (PEN45 in 2009 and PEN79 in 2010). The actual choice for only the PEN45 population was made based on a thorough comparison of both populations. In summary, after the sensory evaluations it became clear that the ranges in which the attributes were scored in the populations were generally more narrow in the PEN79 BIL compared to the PEN45 population (Table 4), indicating less available flavor variation (for mapping). For instance, for the attribute juiciness the difference between the maximum and minimum score in the PEN45 population was 50 (72-22), whereas for the PEN79 material this was only 34 (59-25). This difference between populations becomes even more obvious, when it is realized that the PEN45 sensory data in Table 4 is corrected for taste session and panelist after removal of strong outliers, while for PEN79 the uncorrected data is given. Correction and removal of outliers (extremes) from the PEN79 data would, logically, lead to even further narrowing of the ranges. Also for the interesting attributes aroma intensity and fruity/apple taste the more restricted ranges in the PEN79 population are clear, which is especially caused by lower maximum values of these attributes. This is rather striking as the recurrent parent Vania was scored more sweet and fruity than GNM, while the maximum values of these attributes in the Vania-PEN79 BIL were much lower than in the GNM-PEN45 population (Table 4).

Table 4. Comparison of attribute and non-volatile scores of the PEN45 and PEN79 populations

		PEN45 BIL[1]				PEN79 BIL[2]				Range[3]	
Class	**Trait**	**GNM**	**PEN45**	**Min**	**Max**	**Vania**	**PEN79**	**Min**	**Max**	**PEN45**	**PEN79**
Sensory	Crunchy	64	*	42	79	50	*	31	68	37	37
attributes	Sticky	38	*	17	74	31	*	18	59	57	41
	Tough	23	*	9	72	21	*	11	57	63	46
	Juicy	64	*	22	72	50	*	25	59	50	34
	Sweet	34	*	17	54	47	*	28	65	37	37
	Sour	26	*	6	41	40	*	17	46	35	29
	Aroma	38	*	16	68	48	*	24	58	52	34
	Fruity	10	*	0	41	19	*	0	32	41	32
Biochemical	°Brix	8.3	11.0	6.8	17.9	9.5	15.4	7.7	12.7	11.1	5.0
composition	Glucose	2.8	4.1	1.9	6.4	3.6	4.3	2.8	4.7	4.5	1.9
	Fructose	3.1	3.3	2.2	6.7	3.3	3.5	2.4	4.4	4.5	2.0
	Malate	21.9	80.0	14.1	115.4	28.7	237.9	12.8	41.6	101.3	28.8
	Citrate	364.8	1037.7	210.0	820.0	349.8	1441.8	238.3	450.6	610.0	212.3

[1] Corrected and outlier removed sensory data (copy from Table 1, Chapter 5). [2] Uncorrected (raw) sensory data. [3] Maximum – minimum.

Additionally, during the evaluation of the PEN79 derived material no positive remarks were made by the panel, whereas in the PEN45 taste sessions several positive remarks on the evaluation sheets appeared with notes like ‘presence of tropical fruit flavor’ or ‘similar to papaya taste’. Also the variation of the biochemical compounds in the PEN79 population was much lower than in the PEN45 material, as illustrated for Brix, glucose, fructose, malate and citrate (Table 4). Again this is

notable, as for all these compounds both Vania and PEN79 have higher values than GNM and PEN45, respectively, while the maximum values found in the PEN45 BIL outperform the maxima of the PEN79 material. This comparison between both populations indicates therefore, that the combination of donor and recurrent parent is of utmost importance for a successful project. It also shows, however, that in the case of flavor and biochemical compounds it is not (always) possible to predict on beforehand which combination would give the best offspring. Comparing further, based on the description of the sensory panel and the biochemical measurements from the PEN45 population, we chose three BC2S1 plants with fruits that had either an extraordinary flavor resulting from high sweetness, sourness and/or odor scores or a high sugar and acid concentration (Chapter 5). Using the same criteria for the PEN79 material did not allow us to select a single PEN79 derived plant or line with a similar potential, as any of the three selected PEN45 BC2S1 plants. Finally, also genetic mapping within the PEN79 population turned out to be rather complicated, as only a very poor population specific genetic linkage map was available and the lines were solely genotyped with AFLP markers, that could not be linked to our integrated map. Taking all these arguments into account, led us to the decision only to continue studying the PEN45 derived material, which turned out to be rather challenging already.

Also for PEN45 the development of a solid population specific genetic linkage map proved to be cumbersome, although necessary as genome rearrangements between cultivated and wild species of *Capsicum* are commonly reported (Wu et al. 2009). The main difficulties were the high number of parents (4) of the population in combination with a low number of offspring; 91 BC2 plants, originating from 11 BC1 out of 3 different F1 plants, resulting in a limited amount of effective recombinations. Such a complicated multi-parent population was created, because generation of a single bi-parental population turned out to be impossible due to difficulties in interspecific crossing. In the end however, with a large set of almost 1000 SNP markers, in addition to the 412 AFLP markers that were screened already before the start of the project, we were able to create linkage groups that could be anchored to our integrated map, which corresponds with the numbering and orientation of the chromosomes on the reference maps in the public domain (Barchi et al. 2007, Wu et al. 2009). Although this map allowed us to map several interesting traits (Chapter 5), we realize however, that it is still not perfect as e.g. chromosome 5 is rather underrepresented by one linkage group of only two markers. All in all, it took us about two years to come to the final population specific linkage map (Fig. 2, Chapter 5). This was also the main reason why we performed the QTL mapping on the BC2S1 population (2009 experiment) at the same time as the analysis of

the results from the derived NILs (2011 experiment). A rather unorthodox approach one might think, as development of NILs generally follows after identification of genetic effects. Due to the mentioned limitations, this was however not possible in our case. On the other hand, it also forced us to focus on retention of the phenotypes of interest (rather than gene(tic)s of interest) during NIL development ánd to follow a strategy which would assure maintaining of all (genetic) variation of the initial three BC2S1 plants. Especially, the latter approach did not result in NILs with unique single introgressions, but in a population of NILs with multiple, often overlapping, homozygous ór heterozygous introgressions. The introgressions that were still segregating in the lines, resulted from the time-consuming advancement of the genotyping; i.e. when the NIL experiment was already ended, it was only possible to run the final set of markers on them, which led to the detection of some remaining heterozygous fragments (Fig. 3, Chapter 5). With the knowledge of today and the possibility of re-doing the genotyping, we would start with a (nowadays affordable) RNA or DNA re-sequencing project on the parents for generation of sufficient (>500) genome-wide, polymorphic SNPs, followed by a genotyping project including all BC2 and BC2S1 individuals at once, allowing the creation of a good quality map ánd QTL mapping right after phenotyping.

Several limitations of our BC2S1 mapping population are discussed in Chapter 5 (i.e. accidental co-segregation, underrepresentation of color linked markers and pre-selection leading to skewness), which might have resulted in false positive (as exemplified in Chapter 5) or missed QTLs. An additional limitation or source for missed QTLs might be contributed to the fact that the BC2S1 lines were segregating within the plots they were grown. Although we could manage with this in case of the physical and (some of) the biochemical measurements by making single plant evaluations, for the sensory evaluations, in which we had to combine fruits from multiple plants per plot in one sample for the taste panel, this was not possible. In some cases this led to contradicting scores between panelists, who received different fruit pieces or even led to confusion of individual panelists, that we noticed from remarks, like ‘both sweet and sour fruit pieces present in the same sample’, on the evaluation sheets. Naturally, we tried to avoid this as much as possible, however it occurred occasionally.

Even though the rather complex structure of our BC2S1 mapping population and all aforementioned limitations, we were still fairly well able to map several biochemical, physical and sensory traits, as demonstrated at first for the (monogenic) control traits red color and pungency in the BC2S1 mapping population and in second instance by validation of genetic effects via the NILs. This two-step approach turned out to be very powerful, since it led to the identification of the main results from this

thesis (extensively discussed in Chapter 5): (i) A small *C. baccatum* LG3 introgression causing an extraordinary effect on flavor, which resulted in significantly higher scores for the attributes aroma, flowers, spices, celery and chives. In an attempt to identify the responsible biochemical compounds few consistently up- and down-regulated metabolites were detected. (ii) Two introgressions (LG10.1 and LG1) had major effects on terpenoid content of mature fruits, affecting at least fifteen different monoterpenes. (iii) A second LG3 fragment resulted in a strong increase in Brix (total soluble solids) without negative effects on fruit size. In retrospect, from these three results, the flavor effect caused by the LG3 introgression was the only trait that we actively followed during development of the NILs, by tasting fruits of the different crossing and inbreeding generations needed to create the NILs. The other two results, can be regarded as spin-offs from our followed strategy in maintaining as much as possible *C. baccatum* variation. One could argue therefore as well, that PEN45 might harbor even more variation interesting for *C. annuum* breeding, as we derived our NILs from only three BC2S1 plants.

Concluding remarks

Flavor of pepper is complex and it is influenced by many factors, like the environmental conditions in which the fruits are grown, the interaction between many flavor related metabolites and the flavor perception and/or preferences of consumers. With this thesis we deliver a substantial contribution to the understanding of some of these aspects. We determined the major sensory attributes responsible for the variation in flavor of sweet pepper and identified key-metabolites that influence these attributes. In addition, we developed pre-breeding material with extraordinary flavors outside the scope of varieties currently on the market. In both cases we have shown that metabolite profiling is a valuable ánd reliable tool in breeding for flavor, making targeted improvement of flavor components (attributes) feasible. Once good relations between sensory attributes and metabolites or texture measures have been established, instrumental measurements, either metabolomics approaches or texture analyses, are more attractive than panel evaluations which are costly, low-throughput, time-consuming and rather variable for certain attributes. Still a complicating factor in pepper flavor research is that the effect and interaction of individual attributes on overall consumer liking is not completely clear. Verkerke and Janse (1998), however, reported already 15 years ago, that a ‘good tasting’ pepper should be sweet and crunchy with a fruity aroma. New consumer preference studies are required to confirm this, but it suggests already that compounds like fructose, glucose, (E)-2-hexen-1-ol and p-menth-1-en-9-al, with a positive contribution to both sweetness

and fruity/apple flavor, are interesting candidates to increase the concentration of by breeding. Taking all this together we were able to draw a flowchart for flavor improvement in breeding (Fig. 1).

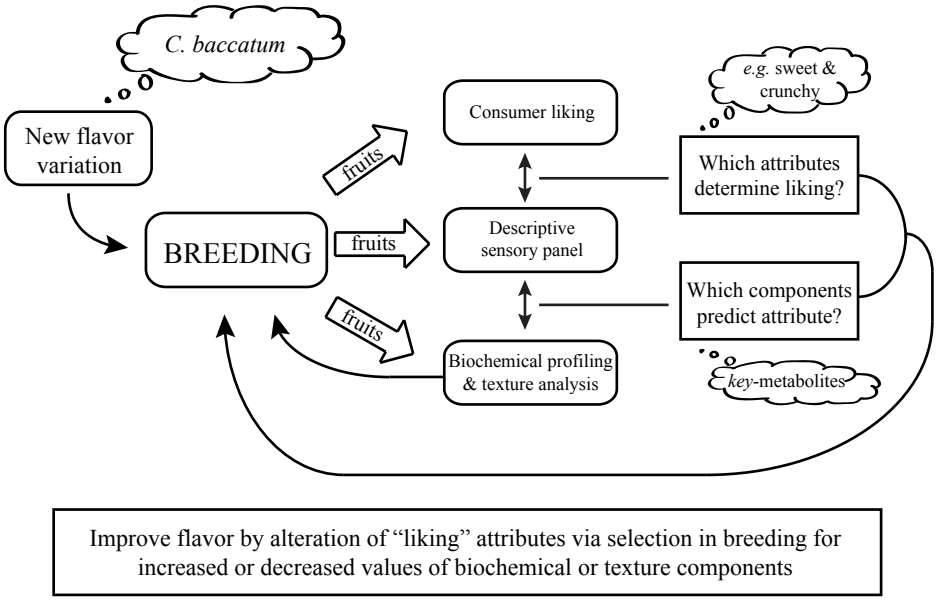


Figure 1. Flowchart for flavor improvement by breeding

In this thesis we also demonstrated that *Capsicum baccatum* is a valuable source for enrichment of the *C. annuum* genetic pool. In several cases unexpected traits were introgressed, as shown by the wide flavor variation, transgressive terpenoid levels and the fruit size unrelated Brix effect. The combination of developed populations allowed mapping of simple morphological traits, but also genetic dissection of quantitative traits with complex inheritance patterns. We realize as well, that we just opened up the flavor wealth wild *Capsicum* species have to offer. In addition, the pile of data we collected over the past 5.5 years, gives ample possibilities for further study. For instance the elite group of genotypes contained three hybrids from which also the parent lines were included, which gives the opportunity to study hybrid effects, similar to what was recently done by Moreno et al. (2012). A proper genotyping of the PEN79 BIL population would be another opportunity, that would directly allow QTL mapping of the available sensory and biochemical data and would make a straight comparison with our PEN45 results feasible. Making use of the current *C. annuum* genome

sequence knowledge in combination with candidate genes from syntenic crops, like tomato, would definitely allow further elucidation of our studied traits. For instance, for the NILs with elevated monoterpene concentrations, clear candidate genes have been identified already. The next step would be to clone the genes (map based or via complementation) and to study their antimicrobial potential. Finally, the availability of the LG3 flavor and Brix NILs, containing a limited amount of introgressions of restricted size, make flavor enhancement in commercial breeding possible today.

References

References

Acharjee A, Kloosterman B, De Vos RCH, Werij JS, Bachem CWB, Visser RGF, Maliepaard C (2011) Data integration and network reconstruction with ~omics data using Random Forest regression in potato. Analytica Chimica Acta 705: 56-63

Alimi NA, Bink MCAM, Dieleman JA, Nicolaï M, Wubs M, Heuvelink E, Magan J, Voorrips RE, Jansen J, Rodrigues PC, van der Heijden GWAM, Vercauteren A, Vuylsteke M, Song Y, Glasbey C, Barocsi A, Lefebvre V, Palloix V, van Eeuwijk FA (2013) Genetic and QTL analyses of yield and a set of physiological traits in pepper. Euphytica 190: 181-201

Arumuganathan K, Earle E (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9: 208-218

Ashrafi H, Hill T, Stoffel K, Kozik A, Yao J, et al. (2012) De novo assembly of the pepper transcriptome (Capsicum annuum): A benchmark for in silico discovery of SNPs, SSRs and candidate genes. BMC Genomics 13: 571

Baldwin EA, Goodner K, Plotto A (2008) Interaction of volatiles, sugars, and acids on perception of tomato aroma and flavor descriptors. Journal of food science 73: 294-307

Baldwin EA, Scott JW, Einstein MA, Malundo TMM, Carr BT, Shewfelt RL, Tandon KS (1998) Relationship between sensory and instrumental analysis for tomato flavour. Journal of the American Society of Horticultural Science 123: 906-915.

Baldwin EA, Scott JW, Shewmaker CK, Schuch W (2000) Flavour trivia and tomato aroma: biochemistry and possible mechanisms for control of important aroma components. Hortscience 35: 1013-1022.

Barchi L, Bonnet J, Boudet C, Signoret P, Nagy I, Lanteri S, Palloix A, Lefebvre V (2007) A high-resolution, intraspecific linkage map of pepper (Capsicum annuum L.) and selection of reduced recombinant inbred line subsets for fast mapping. Genome 50: 51-60

Ben Chaim A, Borovsky Y, Rao GU, Tanyolac B, Paran I (2003) fs3.1: a major fruit shape QTL conserved in Capsicum. Genome 46: 1-9

Black LL, Hobbs HA, Gatti JM (1991) Tomato spotted wilt virus resistance in Capsicum chinense 'PI 152225' and 'PI 159236'. Plant Dis 75: 863

Blum E, Liu K, Mazourek M, Yoo EY, Jahn M, Paran I (2002) Molecular mapping of the C locus for presence of pungency in Capsicum. Genome 45: 702-705

Bohlmann J, Meyer-Gauen G, Croteau R (1998) Plant terpenoid synthases: molecular biology and phylogenetic analysis. Proc Natl Acad Sci USA 95: 4126-4133

Bosland PW, Votava EJ (1999) Peppers: Vegetable and Spice Capsicums. In: Crop Production Science in Horticulture 12. CABI Publishing, Wallingford

Boukema, IW (1982) Resistance to TMV in Capsicum chacoense Hunz. is governed by an allele of the L-locus. Capsicum Newsletter 3:47-48

Brand A, Borovsky Y, Meir S, Rogachev I, Aharoni A, Paran I (2012) pc8.1, a major QTL for pigment content in pepper fruit, is associated with variation in plastid compartment

size. Planta 235: 579-588
Breiman L (2001) Random forests. Machine Learning 45: 5–32
Brewer MS (2009) The effects of irradiation on meat flavor: an overview. Food Engineering and Ingredients 32: 28-30
Bucheli P, Voirol E, de la Torre R, Lopez J, Rytz A, Tanksley SD, Petiard V (1999) Definition of nonvolatile markers for flavor of tomato (Lycopersicon esculentum Mill.) as tools in selection and breeding. J Agric Food Chem 47: 659-664
Buttery RG, Seifert RM, Guadagni DG, Ling LC (1969) Characterization of some volatile constituents of bell peppers. Journal of Agricultural and Food Chemistry 17: 1322-1327
Buttery RG (1993) Quantitative and sensory aspects of flavour of tomato and other vegetables and fruits. In: Acree TE, Teranishsi R (eds) Flavor science: sensible principles and techniques. American Chemical Society, Washington pp 259-286
Buttery RG, Ling LC (1993) Volatile components of tomato fruit and plant parts: relationship and biogenesis. ACS Symposium Series 525: 23-34
Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J. Exp. Bot. 55: 1671-1685
Cookson PJ, Kiano JW, Shipton CA, Fraser PD, Romer S, Schuch W, Bramley PM, Pyke KA (2003) Increases in cell elongation, plastid compartment size and phytoene synthase activity underlie the phenotype of the high pigment-1 mutant of tomato. Planta 217: 896–903
Crosby KM (2008) Pepper. In: Prohens J, Nuez F (eds) Handbook of plant breeding: Vegetables II, Springer, New York, pp 221-248
Do Rêgo ER, Do Rêgo MM, Finger FL, Cruz CD, Casali VWD (2009) A diallel study of yield components and fruit quality in chilli chilli pepper (Capsicum baccatum). Euphytica 168: 275-287
Dunkler D, Sánchez-Cabo F, Heinze G (2011) Statistical analysis principles for Omics data. In: Mayer B (ed) Bioinformatics for Omics Data: Methods and Protocols, Methods in Molecular Biology 719, Springer, New York, pp 113-131
Eggink PM, Haanstra JPW, Tikunov Y, Bovy AG, Visser RGF (2010) Characterization of volatile and non-volatile compounds of fresh pepper (Capsicum annuum). In Advances in genetics and breeding of Capsicum and Eggplant. Editorial de la Universitat Politècnica de València, Spain, pp. 251-259
Eggink PM, Maliepaard C, Tikunov Y, Haanstra JPW, Bovy AG, Visser RGF (2012a). A taste of sweet pepper: Volatile and non-volatile chemical composition of fresh sweet pepper (Capsicum annuum) in relation to sensory evaluation of taste. Food Chemistry 132: 301-310
Eggink PM, Maliepaard C, Tikunov Y, Haanstra JPW, Pohu-Flament LMM, De Wit-Maljaars SC, Willeboordse-Vos F, Bos S, Benning-de Waard C. De Grauw-van Leeuwen PJ, Freymark G, Bovy AG, Visser RGF (2012b) Prediction of sweet pepper (Capsicum annuum) flavor over different harvests. Euphytica 187: 117-131

Eshed Y, Zamir D (1995) An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141: 1147-1162

Falara V, Akhtar TA, Nguyen TTH, Spyropoulou EA, Bleeker PM, Schauvinhold I, Matsuba Y, Bonini ME, Schilmiller AL, Last RL, Schuurink RC, Pichersky E (2011) The tomato terpene synthase gene family. Plant Physiol. 157: 770-789

Fernie AR, Schauer N (2009) Metabolomics-assisted breeding: a viable option for crop improvement? Trends Genet 25: 39-48

Fernie AR, Tadmor Y, Zamir D (2006) Natural genetic variation for improving crop quality. Current Opinion in Plant Biology 9: 196-202

Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nature Reviews Molecular Cell Biology 5: 763-769

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. USA 97: 4718-4723

Fulton TM, Bucheli P, Voirol E, Lopez J, Pétiard V, Tanksley SD (2002) Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four advanced backcross populations of tomato. Euphytica 127: 163-177

Georgelis N, Scott JW, Baldwin EA (2004) Relationship of tomato fruit sugar concentration with physical and chemical traits and linkage of RAPD markers. J Amer Soc Hort Sci 129: 839-845

Gnayfeed MH, Daood HG, Biacs PA, Alcaraz CF (2001) Content of bioactive compounds in pungent spice red pepper (paprika) as affected by ripening and genotype. J. Sci. Food Agric. 81: 1580-1585

Govindarajan VS (1985). Capsicum production, technology, chemistry and quality. Part I. History, botany, cultivation and primary processing. CRC Critical Reviews in Food Science & Nutrition 22: 109-175

Grandillo S, Zamir D, Tanksley SD (1999) Genetic improvement of processing tomatoes: A 20 years perspective. Euphytica 110: 85-97

Hageman JA, Malosetti M, Van Eeuwijk FA (2010). Two-mode clustering of genotype by trait and genotype by environment data. Euphytica: 1-11

Haldane, JBS (1919) The combination of linkage values, and the calculation of distances between the loci of linked factors. Journal of Genetics 8: 299–309

Janse J (1996) Karakterisering van de smaak van verschillende paprikatypen. In: Proefstation voor Bloemisterij en Glasgroente, 19 December

Jarret RL, Baldwin E, Perkins B, Bushway R, Guthrie K (2007) Diversity of fruit quality characteristics in Capsicum frutescens. HortScience 42: 16-19

Kalemba D, Kunicka A (2003) Antibacterial and antifungal properties of essential oils. Current Medicinal Chemistry 10: 813-829

Kappers IF, Hoogerbrugge H, Bouwmeester HJ, Dicke M (2011) Variation in herbivory-

induced volatiles among cucumber (Cucumis sativus L.) varieties has consequences for the attraction of carnivorous natural enemies. J Chem Ecol 37: 150-160

Keurentjes JJB, Fu J, de Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nature Genetics 38: 842-849

Kim S, Kim KT, Kim DH, Yang EY, Cho MC, Jamal A, Chae Y, Pae DH, Oh DG, Hwang JK (2010) Identification of quantitative trait loci associated with anthracnose resistance in chili pepper (Capsicum spp.). Korean Journal of Horticultural Science and Technology 28: 1014-1024

Kollmannsberger H, Rodriguez-Burruezo A, Nitz S, Nuez F (2007) Comparative analysis of volatile compounds involved in the flavor of Capsicum annuum fruits. In: Niemirowicz-Szczytt K (ed) Progress in Research on Capsicum & Eggplant. Warsaw University of Life Sciences Press, Warsaw, pp 195-203

Kollmannsberger H, Rodríguez-Burruezo A, Nitz S, Nuez F (2011) Volatile and capsaicinoid composition of ají (Capsicum baccatum) and rocoto (Capsicum pubescens), two Andean species of chile peppers. J. Sci. Food Agric. 91: 1598-1611

Kosambi DD (1944) The estimation of map distance from recombination values. Annals of Eugenics 12: 172-175

Krammer GE, Werkhoff P, Sommer H, Schmidt CO, Gatfield I, Bertram HJ (2001) Carotenoid degradation products in paprika powder. In: Winterhalter P et al. (eds) Carotenoid-derived aroma compounds. ACS Symposium Series. Washington, DC, Vol. 802,chapter 15, pp 206-219

Krumbein A, Auerswald H (1998) Characterization of aroma volatiles in tomatoes by sensory analyses. Nahrung 42: 395-399

Kumar R, Dwivedi N, Kumar Singh R, Kumar S, Prakash Rai V, Singh M (2011) A review on molecular characterization of pepper for capsaicin and oleoresin. International Journal of Plant Breeding and Genetics 5: 99-110

Lecomte L, Saliba-Colombani V, Gautier A, Gomez-Jimenez MC, Duffe MC, Buret M, Causse M (2004) Fine mapping of QTLs affecting the fruit architecture and composition of tomato. Mol. Breed. 13: 1-14

Lefebvre V, Daubeze AM, Rouppe van der Voort J, Peleman J, Bardin M, Palloix A (2003) QTLs for resistance to powdery mildew in pepper under natural and artificial infections. Theoretical and Applied Genetics 107: 661–666

Lefebvre V, Kuntz M, Camara B, Palloix A (1998) The capsanthin-capsorubin synthase gene: a candidate gene for the y locus controlling the red fruit colour in pepper. Plant Molecular Biology 36: 785–789

Lewinsohn E, Sitrit Y, Bar E, Azulay Y, Ibdah M, Meir A, Yosef E, Zamir D,Tadmor Y (2005) Not just colors - carotenoid degradation as a link between pigmentation and aroma in tomato and watermelon fruit. Trends in Food Science and Technology 16: 407-415

Luning PA, de Rijk T, Wichers HJ, Roozen JP (1994a) Gas chromatography, mass spectrometry, and sniffing port analyses of volatile compounds of fresh bell peppers

(Capsicum annuum) at different ripening stages. Journal of Agricultural and Food Chemistry 42: 977-983

Luning PA, van der Vuurst de Vries R, Yuksel D, Ebbenhorst-Seller T, Wichers HJ, Roozen JP (1994b) Combined instrumental and sensory evaluation of flavor of fresh bell peppers (Capsicum annuum) harvested at three maturation stages. Journal of Agricultural and Food Chemistry 42: 2855-2861

Meckelmann SW, Riegel DW, van Zonneveld MJ, Ríos L, Peña K, Ugas R, Quinonez L, Mueller-Seitz E, Petz M (2013) Compositional characterization of native Peruvian chili peppers (Capsicum spp.). J Agric Food Chem. 61: 2530-2537

Mihaleva VV, Verhoeven HA, De Vos RCH, Hall RD, Van Ham RCHJ (2009). Automated procedure for candidate compound selection in GC-MS metabolomics based on prediction of Kovats retention index. BioInformatics 25: 787-794

Minguez-Mosquera MI (1998) Color quality in paprika oleoresins. J.Agric. Food Chem. 46: 5124-5127

Moreno E, Fita A, Gonzalez-Mas MC, Rodriguez-Burruezo A (2012) HS-SPME study of the volatile fraction of Capsicum accessions and hybrids in different parts of the fruit. Scientia Horticulturae 135: 87-97

Moretti CL, Sargent SA, Huber DJ, Calbo AG (1998). Chemical composition and physical properties of pericarp locule and placental tissues of tomatoes with internal bruising. Journal of the American Society for Horticultural Science 123: 656-660

Mosciano G, Fasano M, Cassidy J (1993). Organoleptic characteristics of flavor materials. Perfumer and Flavorist 18: 43-45

Mosciano G (1998). Organoleptic characteristics of flavor materials. Perfumer and Flavorist 23: 33-36

Mosciano G (1999). Organoleptic characteristics of flavor materials. Perfumer and Flavorist 24: 49-51

Moscone EA, Baranyi M, Ebert I, Greilhuber J, Ehrendorfer F, et al. (2003) Analysis of nuclear DNA content in Capsicum (Solanaceae) by flow cytometry and Feulgen densitometry. Annals of Botany 92: 21-29

Moscone EA, Scaldaferro MA, Gabriele M, Cecchini NM, Sánchez García Y, Jarret R, Daviña JR, Ducasse DA, Barboza GE, Ehrendorfer F (2007) The evolution in chili peppers (Capsicum-Solanaceae), a cytogenetic perspective. Acta Horticulturae 745: 137-170

Nagegowda DA (2010) Plant volatile terpenoid metabolism: biosynthetic genes, transcriptional regulation and subcellular compartmentation. FEBS letters 584: 2965-2973

Nicolaï M, Pisani C, Bouchet J-P, Vuylsteke M, Palloix A (2012) Discovery of a large set of SNP and SSR genetic markers by high-throughput sequencing of pepper (Capsicum annuum) Genetics and Molecular Research 11: 2295-2300

Onus AN, Pickersgill B (2004) Unilateral incompatibility in Capsicum (Solanaceae): occurrence and taxonomic distribution. Annals of Botany 94: 289-295

Ortiz R, Delgado de la Flor F, Alvarado G, Crossa J (2010) Classifying vegetable genetic resources. A case study with domesticated Capsicum spp. Sci. Hortic. 126: 186-191

Pérez-Sánchez R, Infante F, Gálvez C, Ubera JL (2007) Fungitoxic activity against phytopathogenic fungi and the chemical composition of Thymus zygis essential oils. Food Sci Tech Int 13: 341-347

Perry L, Dickau R, Zarrillo S, Holst I, Pearsall DM, Piperno DR, Berman MJ, Cooke RG, Rademaker K, Ranere AJ, Raymond JS, Sandweiss DH, Scaramelli F, Tarble K, Zeidler JA (2007) Starch fossils and the domestication and dispersal of chili peppers (Capsicum spp. L.) in the Americas. Science 315: 986-988

Pétro-Turza M (1986) Flavor of tomato and tomato products. Food Rev. Int. 2: 309-351

Pickersgill B (1997) Genetic resources and breeding of Capsicum spp. Euphytica 96: 129-133

Poulos J (1994) Pepper breeding (Capsicum spp.): achievements, challenges and possibilities. Plant Breeding Abstracts 64: 143-155

Stevens MA, Kader AA, Albright-Holton M, Algazi M (1977) Genotypic variation for flavor and composition in fresh market tomatoes. J Am Soc Hort Sci 102: 680-689

Stewart C, Kang BC, Liu K, Mazourek M, Moore SL, Yoo EY, Kim BD, Paran I, Jahn MM (2005) The Pun1 gene for pungency in pepper encodes a putative acyltransferase. Plant Journal 42: 675–688

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5: 94-100

Ritota M, Marini F, Sequi P, Valentini M (2010) Metabolomic characterization of Italian sweet pepper (Capsicum annuum L.) by means of HRMAS-NMR spectroscopy and multivariate analysis. J Agric Food Chem. 58: 9675-9684

Rocha MC, Deliza R, Corrêa FM, do Carmo MGF, Abboud ACS (2013) A study to guide breeding of new cultivars of organic cherry tomato following a consumer-driven approach. Food Research International 51: 265-273

Rodriguez-Burruezo A, Kollmannsberger H, Gonzalez-Mas MC, Nitz S, Nuez F (2010). HS-SPME comparative analysis of genotypic diversity in volatile fraction and aroma contributing compounds of Capsicum fruits from the annuum-chinense-frutescens complex. Journal of Agricultural and Food Chemistry 58: 4388-4400

Rodriguez-Burruezo A, Prohens J, Raigon MD, Nuez F (2009) Variation for bioactive compounds in aji (Capsicum baccatum L.) and rocoto (C. pubescens R. & P.) and implications for breeding. Euphytica 170:169–181

Ruiz JJ, Alonso A, Garcia-Martinez S, Valero M, Blasco P, Ruiz-Bevia F (2005) Quantitative analysis of flavour volatiles detects differences among closely related traditional cultivars of tomato. Journal of the Science of Food and Agriculture 85: 54-60

Sinesio F, Cammareri M, Moneta E, Navez B, Peparaio M, Causse M, Grandillo S (2010) Sensory quality of fresh French and Dutch market tomatoes: a preference mapping study with Italian consumers. Journal of Food Science 75: 55-67

Tadmor Y, Fridman E, Gur A, Larkov O, Lastochkin E, Ravid U, Zamir D, Lewinsohn E (2002) Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. J. Agric. Food Chem. 50: 2005-2009

Tandon KS, Baldwin EA, Scott JW, Shewfelt RL (2003). Linking sensory descriptors to

volatile and non-volatile components of fresh tomato flavor. Journal of Food Science 68: 2366-2371

Tandon KS, Baldwin EA, Shewfelt RL (2000). Aroma perception of individual volatile compounds in fresh tomatoes (Lycopersicon esculentum Mill.) as affected by the medium of evaluation. Postharvest Biology and Technology 20: 261-268

Thorup TA, Tanyolac B, Livingstone KD, Popovsky S, Paran I, Jahn M (2000) Candidate gene analysis of organ pigmentation loci in the Solanaceae. PNAS 97: 11192-11197

Tibshirani, R (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B 58: 267-288

Tieman DM, Zeigler M, Schmelz EA, Taylor MG, Bliss P, Kirst M, Klee HJ (2006) Identification of loci affecting flavor volatile emissions in tomato fruits. Journal of Experimental Botany 57: 887-896

Tikunov YM, Laptenok S, Hall RD, Bovy AG, de Vos RCH (2012) MSClust: A tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics 8: 714-718

Tikunov Y, Lommen A, de Vos CHR, Verhoeven HA, Bino RJ, Hall RD, Bovy AG (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. Plant Physiology 139: 1125-1137

Tikunov YM, de Vos RCH, Gonzalez Paramas AM, Hall RD, Bovy AG (2010) A role for differential glycoconjugation in the emission of phenylpropanoid volatiles from tomato fruit discovered using a metabolic data fusion approach. Plant Physiology 152: 55-70

Vallverdu-Queralt A, Bendini A, Tesini F, Valli E, Lamuela-Raventos RM, Toschi TG (2013) Chemical and sensory analysis of commercial tomato juices present on the Italian and Spanish markets. Journal of agricultural and food chemistry 61: 1044-1050

Van Ooijen JW (2006) JoinMap 4.0: software for the calculation of genetic linkage maps in experimental populations. Plant Research International, Wageningen, The Netherlands

Van Ooijen JW (2009) MapQTL 6: software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV, Wageningen, The Netherlands

Van Ruth SM, Roozen JP (1994) Gas chromatography/sniffing port analysis and sensory evaluation of commercially dried bell peppers (Capsicum annuum) after rehydration. Food Chemistry 51: 165-170

Van Ruth SM, Roozen JP, Cozijnsen JL, Posthumus MA (1995) Volatile compounds of rehydrated French beans, bell peppers and leeks. Part II. Gas chromatography/sniffing port analysis and sensory evaluation. Food Chemistry 54: 1-7

Velterop JS, Vos F (2001). A rapid and inexpensive microplate assay for the enzymatic determination of glucose, fructose, sucrose, L-malate and citrate in tomato (Lycopersicon esculentum) extracts and in orange juice. Phytochemical analysis 12: 299-304

Verheul J (2008) WUR: smaak blijft achter. Agrarisch Dagblad, 18 September

Verkerke W, Janse J (1998). Lekkere paprika is zoet zonder bijsmaak. Groenten en fruit, 20 February, p. 6-7

Verkerke W (2000). Paprika smaak kan nog veel beter. Groenten en fruit, 14 January, p. 21

Verkerke W, Kersten M (2000) The role of fruit texture in flavour of tomato and sweet pepper. In: Spatz HC, Speck T (eds) Proceedings 3rd Plant Biomechanics Conference Freiburg. Thieme, Stuttgart, pp 342-347

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93:77–78

Wahyuni Y, Ballester A-R, Sudarmonowati E, Bino RJ, Bovy AG (2013a) Secondary metabolites of Capsicum species and their importance in the human diet. Journal of natural products 76: 783-793

Wahyuni Y, Ballester A-R, Tikunov Y, de Vos CHR, Pelgrom KTB, Maharijaya A, Sudarmonowati E, Bino RJ, Bovy AG (2013b) Metabolomics and molecular marker analysis to explore pepper (Capsicum spp.) biodiversity. Metabolomics 9:130-144

Walling LL (2000) The myriad plant responses to herbivores. Journal of Plant Growth Regulation 19: 195-216

Walsh BM, Hoot SB (2001) Phylogenetic relationships of Capsicum (Solanaceae) using DNA sequences from two noncoding regions: The chloroplast atpB-rbcL spacer region and nuclearwaxy introns. International Journal of Plant Sciences 162: 1409-1418

Wu F, Eannetta NT, Xu Y, Durrett R, Mazourek M, Jahn MM, Tanksley SD (2009) A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus Capsicum. Theor Appl Genet 118: 1279-1293

Yarnes SC, Ashrafi H, Reyes-Chin-Wo S, Hill TA, Stoffel KM, Van Deynze A (2013) Identification of QTLs for capsaicinoids, fruit quality, and plant architecture-related traits in an interspecific Capsicum RIL population. Genome 56: 61-74

Yoon JB, Yang DC, Lee WP, Ahn SY, Park HG (2004) Genetic resources resistant to anthracnose in the genus Capsicum. J. Kor. Soc. Hort. Sci. 45: 318-323

Yoon JB, Yang DC, Wahng Do J, Park HG (2006) Overcoming two post-fertilization genetic barriers in interspecific hybridization between Capsicum annuum and C. baccatum for introgression of anthracnose resistance. Breeding Science 56: 31-38

SUMMARIES

Summary

This PhD project started with the composition of a diverse panel of genotypes that represented, (i) roughly the flavor variation in the commercial *Capsicum annuum* breeding program of Rijk Zwaan, (ii) parents of available mapping populations and (iii) some genotypes that were expected to have extraordinary flavors. The complete set consisted of 35 genotypes of which 24 genotypes were non-pungent. Volatile and non-volatile compounds as well as some breeding parameters were measured in mature fruits of all genotypes throughout the growing season. In addition, from three harvests the non-pungent genotypes were evaluated for taste by a trained descriptive sensory panel.

The biochemical profiling with use of SPME-GC-MS allowed visualization of between- and within-species volatile compound variation. Principal components analysis (PCA) on the intensity patterns of 391 putative volatile compounds revealed individual grouping of *C. chinense, C. baccatum* var. *pendulum* and *C. annuum*, indicating potentially interesting volatile variation present in the former two groups. A large group of saturated and unsaturated esters were mainly responsible for the individual grouping of the *C. chinense* accessions. Due to the elevated acid concentrations and aberrant volatile profiles of the *C. baccatum* var. *pendulum* accessions PEN45 and PEN79, the two BIL populations derived from these accessions were identified as interesting candidates for further study. Compared to e.g. Mazurka the citrate concentration of the *C. baccatum* accessions was 2.5-3 times higher and the malate concentrations were even up to 12 times higher (Chapter 2).

Based on the non-pungent genotypes, we found highly correlated clusters of volatiles and non-volatiles, which could be related to metabolic pathways and common biochemical precursors (Chapter 3). Contrasts between genotypes were caused by both qualitative and quantitative differences in these metabolic clusters, with the phenolic derivatives, higher alkanes, sesquiterpenes and lipid derived volatiles forming the major determinants. For the description of the non-pungent genotypes the panelists used fourteen attributes to describe the flavor sensation in the mouth/throat, which were the texture attributes crunchiness, stickiness of the skin, toughness and juiciness, the basic taste attributes sweetness and sourness and the retronasal flavor attributes aroma intensity, grassiness, green bean, carrot, fruity/apple, perfume, petrochemical and musty. The variation in flavor could be reduced into two major sensory contrasts, which were a texture related contrast and the basic sweet-sour contrast. The structure of the PCA plots resulting from the analysis with one harvest (Chapter 3) and the analysis with the combined three harvests (Chapter 4) remained almost identical, indicating the stability of these contrasts. To relate

the sensory attributes to the metabolite data and to determine the importance of the individual compounds we used Random Forest regression on the individual harvests and on the three harvests together. Several predictors for the attributes aroma, fruity/apple, sourness and sweetness were found in common between harvests, which we proposed as key-metabolites involved in flavor determination of sweet pepper (Chapter 4). This list contains compounds with known relations to attributes, like sweetness and sugars, but also several compounds with new relations. In this respect we have demonstrated for the first time, that the metabolites p-menth-1-en-9-al, (E)-β-ocimene, (Z)-2-penten-1-ol, and 1-methyl-1,4-cyclohexadiene are related to fruity/apple taste and/or sweetness of pepper. For sourness the only compound with a consistent significant contribution was an unknown $C_6H_8O_2$ compound. We postulated therefore the hypothesis that in pepper the effect of sourness related metabolites is masked by other volatile and non-volatile compounds or texture differences (Chapter 3). Subsequently in Chapter 4 we described a clear sweetness-sourness interaction and demonstrated that the masking effect of fructose and other sugars explained why we did not find organic acids contributing to the prediction of sourness. The major sensory attributes were also predicted between harvests. The Random Forest predictions of the texture related attributes (juiciness, toughness, crunchiness and stickiness of the skin) and sweetness were very good. The predictions of the attributes aroma intensity, sourness and fruity/apple were somewhat lower and more variable between harvests, especially in the second harvest. In general, we concluded that prediction of attributes with higher heritabilities works better and is more consistent over harvests (Chapter 4).

Based on the results of the initial experiments (Chapter 2) the species *C. baccatum* was chosen for further study. To exploit the potential flavor wealth of *C. baccatum* PEN45 we combined interspecific crossing with embryo rescue, resulting in a multi-parent BC2S1 population, that was characterized for sensory and biochemical variation (Chapter 5). We developed a population specific genetic linkage map for QTL mapping of characterized traits. Because of the complex structure of our BC2S1 mapping population we encountered several limitations, such as accidental co-segregation, underrepresentation of color linked markers and pre-selection leading to skewness, which might have resulted in false positive or missed QTLs. Despite these limitations, we were still fairly well able to map several biochemical, physical and sensory traits, as demonstrated at first for the (monogenic) control traits red color and pungency in the BC2S1 mapping population and in second instance by validation of genetic effects via an experiment with near-isogenic lines (NILs). This two-step approach turned out to be very powerful, since it led to the identification of the main results from this thesis: (i) A small *C. baccatum* LG3 introgression

causing an extraordinary effect on flavor, which resulted in significantly higher scores for the attributes aroma, flowers, spices, celery and chives. In an attempt to identify the responsible biochemical compounds few consistently up- and down-regulated metabolites were detected, including the well-known pepper compound 2-isobutyl-3-methoxypyrazine (down) and 6-methyl-4-oxo-5-heptenal (up); (ii) Two introgressions (LG10.1 and LG1) had major effects on terpenoid content of mature fruits, affecting at least fifteen different monoterpenes; (iii) A second LG3 fragment resulted in a strong increase in Brix (total soluble solids) without negative effects on fruit size (Chapter 5).

In Chapter 6 some extra sensory results of the pungent genotypes are given and a comparison between the two *C. baccatum pendulum* BILs (PEN45 and PEN79 derived) is made in light of the overall results. Finally the perspectives for breeding are discussed and presented in the form of a flowchart for flavor improvement.

SAMENVATTING

Dit promotieonderzoek is gestart met het samenstellen van een set genotypen met grote diversiteit die (i) grofweg de variatie in smaak binnen het commerciële *Capsicum annuum* veredelingsprogramma van Rijk Zwaan vertegenwoordigt, (ii) ouders van beschikbare karteringspopulaties bevat en (iii) enkele genotypen omvat, waar een buitengewone smaak van werd verwacht. De complete set bestond uit 35 genotypen, waarvan er 24 niet-heet waren. Vluchtige en niet-vluchtige stoffen, aangevuld met enkele veredelingscriteria werden gedurende de teelt gemeten in rijpe vruchten van alle genotypen. Daarnaast zijn van drie oogsten de niet-hete genotypen beoordeeld op smaak door een getraind beschrijvend sensorisch panel.

De biochemische karakterisatie met behulp van SPME-GC-MS heeft de vluchtige stoffen variatie binnen en tussen de *Capsicum* soorten zichtbaar gemaakt. Principale componenten analyse (PCA) op basis van intensiteiten van 391 potentiële vluchtige stoffen onthulde individuele groepen met *C. chinense, C. baccatum* var. *pendulum* en *C. annuum* accessies, hetgeen wijst op de aanwezigheid van mogelijk interessante vluchtige stof variatie in de eerste twee groepen. Een grote groep verzadigde en onverzadigde esters was voornamelijk verantwoordelijk voor de afzonderlijke groepering van de *C. chinense* accessies. Vanwege de verhoogde zuur concentraties en afwijkende vluchtige stof profielen van de *C. baccatum* var. *pendulum* accessies PEN45 en PEN79, zijn de twee geavanceerde terugkruisingspopulaties (backcross inbred lines; BIL) afgeleid van deze accessies geïdentificeerd als interessante kandidaten voor vervolgstudie. De citraat concentratie van de *C. baccatum* accessies was ten opzichte van Mazurka 2.5-3 keer hoger; de malaat concentraties waren zelfs tot twaalf keer hoger (Hoofdstuk 2).

Op basis van de niet-hete genotypen hebben we sterk gecorreleerde groepen met vluchtige en niet-vluchtige stoffen gevonden, die gerelateerd konden worden aan gemeenschappelijke biochemische herkomst (Hoofdstuk 3). Contrasten tussen genotypen werden veroorzaakt door zowel kwalitatieve als kwantitatieve verschillen in deze groepen met metabolieten, waarbij de fenolische derivaten, hogere alkanen, sesquiterpenen en vetzuur afgeleide vluchtige stoffen de voornaamste determinanten waren. Voor de beschrijving van de niet-hete genotypen gebruikten de smaakpanelleden veertien attributen om de smaaksensatie in de mond en keel te typeren. Dit waren de textuurattributen knapperigheid, nahangen van de schil, taaiheid en sappigheid, de basissmaken zoetheid en zuurheid en de retronasale attributen aroma intensiteit, grassig, snijbonen, peen, fruitig, parfum, petrochemisch en muf. De smaakvariatie kon worden teruggebracht tot de twee voornaamste sensorische contrasten; een textuur gerelateerd contrast en het basale zoet-zuur contrast. De

structuur van de PCA plots vanuit de analyse met één oogst (Hoofdstuk 3) en de analyse met de drie oogsten gezamenlijk (Hoofdstuk 4) bleef bijna identiek, wat de stabiliteit van deze contrasten duidelijk weergeeft. Om de sensorische attributen te kunnen relateren aan de metabolieten en het belang van individuele stoffen te kunnen bepalen hebben we Random Forest regressie toegepast op de individuele oogsten en de drie oogsten gezamenlijk. Voor de attributen aroma, fruitigheid, zuurheid en zoetheid werden verschillende gemeenschappelijke voorspellers gevonden tussen de oogsten, welke we hebben voorgesteld als de sleutel-metabolieten verantwoordelijk voor de smaakbepaling van paprika (Hoofdstuk 4). Deze lijst bevat stoffen met bekende relaties tot attributen, zoals zoetheid en suikers, maar ook verschillende stoffen met nieuwe relaties. In dit opzicht hebben we voor het eerst aangetoond dat de metabolieten p-menth-1-en-9-al, (E)-β-ocimeen, (Z)-2-penten-1-ol, en 1-methyl-1,4-cyclohexadieen gerelateerd zijn aan fruitige smaak en/of zoetheid van paprika. Voor zuurheid was de enige stof met een consistent significante bijdrage een onbekende $C_6H_8O_2$ stof. We hebben daarom de hypothese opgesteld dat in paprika het effect van zuurheid gerelateerde metabolieten wordt gemaskeerd door andere vluchtige en niet-vluchtige stoffen of textuur verschillen (Hoofdstuk 3). Vervolgens hebben we in Hoofdstuk 4 een duidelijke zoet-zuur interactie beschreven en aangetoond dat het maskerende effect van fructose en andere suikers verklaart waarom we geen organische zuren met bijdrage aan de voorspelling van zuurheid hebben gevonden. De belangrijkste sensorische attributen hebben we ook voorspeld tussen de oogsten. De Random Forest voorspellingen van de textuur gerelateerde attributen (sappigheid, taaiheid, knapperigheid en nahangen van de schil) en zoetheid waren erg goed. De voorspellingen van de attributen aroma intensiteit, zuurheid en fruitig waren wat minder goed en meer variabel tussen oogsten, met name in de tweede oogst. In het algemeen hebben we geconcludeerd dat voorspelling van attributen met hogere vererfbaarheid beter werkt en consistenter is tussen de oogsten (Hoofdstuk 4).

Op basis van de resultaten van de initiële experimenten (Hoofdstuk 2) werd de soort *C. baccatum* gekozen voor verdere bestudering. Om de potentiële overvloed aan smaak variatie vanuit *C. baccatum* PEN45 te kunnen onderzoeken, hebben we interspecifieke kruisingen gemaakt in combinatie met embryo rescue, wat resulteerde in een multi-ouder BC2S1 populatie, die werd gekarakteriseerd voor sensorische en biochemische variatie (Chapter 5). Ten behoeve van QTL kartering van de gekarakteriseerde eigenschappen hebben we een populatie specifieke genetische koppelingskaart gemaakt. Door de complexe structuur van onze BC2S1 karteringspopulatie hebben we verschillende beperkingen ondervonden, zoals accidentele co-uitsplitsing, ondervertegenwoordiging van merkers gekoppeld aan kleur en voorselectie leidend tot scheefheid, wat geresulteerd kan hebben in vals-

positieve of gemiste QTLs. Ondanks deze beperkingen zijn we nog steeds behoorlijk goed in staat geweest om verschillende biochemische, fysieke en sensorische eigenschappen te karteren, zoals in eerste instantie getoond voor de (monogene) controle eigenschappen rode kleur en heetheid in de BC2S1 populatie, en in tweede instantie door validatie van genetische effecten met behulp van een experiment met nabij-isogene lijnen (near isogenic Lines; NILs). Deze twee-traps aanpak bleek erg krachtig te zijn, aangezien het leidde tot de identificatie van de belangrijkste resultaten van dit proefschrift: (i) Een kleine *C. baccatum* LG3 introgressie met een buitengewoon effect op smaak, resulterend in significant hogere scores voor de attributen aroma, bloemen, kruidig, selderij en peterselie. In een poging om de verantwoordelijke biochemische stoffen te identificeren, hebben we verschillende consistent omhoog- en omlaag gereguleerde metabolieten gevonden, inclusief de welbekende paprika stof 2-isobutyl-3-methoxypyrazine (omlaag) and 6-methyl-4-oxo-5-heptenal (omhoog); (ii) Twee introgressies (LG10.1 en LG1) hadden een zeer groot effect op de gehaltes terpenoïden in de rijpe vruchten, met een effect op minstens vijftien verschillende monoterpenen; (iii) Een tweede LG3 fragment zorgde voor een sterke toename in Brix (hoeveelheid opgeloste stof) zonder nadelige effecten op vruchtgrootte (Hoofdstuk 5).

In Hoofdstuk 6 zijn nog enkele extra sensorische resultaten van de hete genotypen weer gegeven en is er een vergelijking tussen de twee *C. baccatum pendulum* BILs (afgeleid van PEN45 en PEN79) gemaakt in het licht van de algehele resultaten. Tenslotte worden de perspectieven voor veredeling bediscussieerd en gepresenteerd in de vorm van een flowchart voor smaak verbetering.

Dankwoord
Curriculum vitae

Een groot deel van mijn jeugd heb ik in mijn vrije tijd in de bloembollen gewerkt. Na het afronden van het VWO in 2001 was het voor mij dan ook een logische keuze om naar Wageningen te gaan om Plant en Gewaswetenschappen te gaan studeren om uiteindelijk bloembollenveredelaar te kunnen worden. De veredeling bleef trekken tijdens de studie, maar het vooruitzicht dat ik 6-7 jaar zou moeten wachten op het resultaat van een kruising; dit is de gemiddelde duur bij tulp van een kruising tot weer een bloeiende bol, deed er toe leiden dat ik ook eens bij andere gewassen mijn licht ging opsteken. Zo kwam ik aan het einde van het derde jaar via Fred van Eeuwijk met Johan Schut van Rijk Zwaan in contact om een karteringsproject in sla uit te voeren, waarover ik een rapport schreef ter afronding van de BSc. In sla bleek de generatieduur al een stuk korter te zijn en ik leerde dat er binnen een bedrijf als Rijk Zwaan ook allerlei high-tech onderzoeksprojecten werden uitgevoerd, vergelijkbaar met projecten bij de universiteit. Een hele openbaring voor mij in die tijd.

Tijdens de Master fase van de studie heb ik eerst een minor thesis bij Fytopathologie gedaan gericht op karakterisatie van NEP eiwitten (necrosis and ethylene-inducing proteins) van het pathogeen *Botrytis* in diverse plantensoorten. Van mijn begeleider Sander Schouten heb ik toen wel geleerd hoe leuk onderzoek eigenlijk is. Vervolgens heb ik mijn major thesis uitgevoerd bij Plantenveredeling met als onderwerp zaadloosheid (parthenocarpy) in tomaat. Een intrigerend onderwerp (dat nog regelmatig in gesprekken met vrienden terugkomt…) met name omdat de natuur er normaal gesproken op gericht is om juist voldoende zaad te vormen om nakomelingschap te garanderen. De combinatie van fenotypering in de kas, moleculair werk in het lab en de analyse van gegenereerde data vond ik prachtig. Eindelijk een project waar ik praktisch aan de slag kon, maar met meer dan voldoende intellectuele uitdaging. Begeleiders Benoit Gorguet en Pim Lindhout zagen dit ook en wisten mij er steeds meer van te overtuigen dat een promotieonderzoek echt iets voor mij zou zijn. Dank daarvoor! Tot die tijd was ik er namelijk van overtuigd dat ik absoluut geen PhD zou gaan doen.

Complicerende factor was echter dat ik op dat moment al wel door Rijk Zwaan was benaderd om te solliciteren op de baan van Pre-breeder *Solanaceae*. Een uitgelezen kans voor mij om toegepast onderzoek te gaan doen in de gewassen tomaat, paprika en aubergine met een generatietijd van ‘maar’ een half jaar en de beschikbaarheid van de nodige moleculaire tools. Het kan haast niet anders dan dat Manja Verhoef toen een goed woordje voor me heeft gedaan bij haar vader, die bij Rijk Zwaan werkt. Anders weet ik niet of ik er zo makkelijk zou zijn ingerold. In ieder geval, nog bijna een jaar voor de afronding van mijn studie had ik ook eigenlijk niets te verliezen en ben ik het sollicitatietraject in gegaan, waarbij ik op een zeker moment heb gepolst of de baan niet gecombineerd zou kunnen worden met een promotieonderzoek. Ruud

Verhoef (tomaat) en Jair Haanstra (paprika en aubergine) waren direct enthousiast en nadat 'big boss' Kees Reinink ook zijn fiat had gegeven, ben ik aangenomen met het vooruitzicht om de job te combineren met een PhD, waaraan ik 30% van mijn tijd zou mogen besteden. Deze drie heren wil ik dan ook hartelijk bedanken voor het op dat moment al gegeven vertrouwen.

Het eerste jaar (2007) bij Rijk Zwaan hebben we er voor uitgetrokken om enerzijds te kunnen inwerken en anderzijds een geschikt onderwerp te bepalen waarop het promotieonderzoek zich zou kunnen richten. De keuze op het gewas paprika was al snel gevallen aangezien Jair, als gepromoveerd Wageninger, goed de dagelijkse begeleiding kon geven. Bij de keuze voor het onderwerp 'smaak' hebben factoren als, interesse voor Rijk Zwaan, academische diepgang, publiceerbaarheid en concurrentie in het veld meegespeeld. Op dat moment hadden we al contact opgenomen met Sjaak van Heusden (WUR) om te zien wat de mogelijkheden bij de leerstoelgroep Plantenveredeling waren. Sjaak had al vaker een vergelijkbare situatie meegemaakt (o.a. PhD project van Benoit betaald door Western Seeds) en speelde ook nu weer een belangrijke rol in de totstandkoming van mijn PhD project.

In samenspraak met prof. Richard Visser, Sjaak en Rijk Zwaan schreef ik een projectvoorstel dat goedgekeurd werd door de onderzoeksschool EPS (Experimental Plant Sciences). Toen was het zover, op 1 januari 2008 startte mijn PhD officieel. Om een bliksemstart te kunnen maken stonden ondertussen de planten voor het eerste experiment al op de opkweekvloer. De jaren die volgden typeerden zich enerzijds door een steeds verdere ontwikkeling van het nieuwe werkveld pre-breeding binnen Rijk Zwaan, wat zich uitte in een grote toename van het aantal onderzoeksprojecten die ik onder mijn hoede kreeg en typeerden zich anderzijds door de continue noodzaak voor voldoende voortgang binnen mijn promotieonderzoek. Dat dit er toe leidde dat ik dit proefschrift niet binnen vier jaar heb afgerond zal weinig mensen verbazen, evenmin dat sommige mensen eraan hebben getwijfeld of het überhaupt af zou komen. Met name dit laatste was voor mij een extra stimulans om het project wel degelijk af te ronden. Van huis uit kreeg ik al de boodschap mee dat als je ergens aan begint je dit ook afmaakt: *'als je ergens lid van bent...'*. Dit principe heeft me zeker geholpen om niet op te geven.

Naast de mensen die ik eerder al noemde zijn er nog veel, heel veel mensen die ik wil bedanken voor de totstandkoming van dit proefschrift.

Bij PRI (Plant Research International) wil ik Harry Jonker en Yvonne Birnbaum van harte bedanken voor het uitvoeren van de SPME-GC-MS metingen. Ik weet dat het met name in de eerste twee jaar niet makkelijk moet zijn geweest om de

ultra-hete pepers te vermalen. Een zuurkast en gasmasker moesten eraan te pas komen om de capsaicine dampen draaglijk te maken. Weet wel, dat de resultaten van deze hete rakkers uiteindelijk hebben geleid tot de paprika met exotische smaak die we hebben gemaakt binnen dit project. In the same group I would also like to thank Yury Tikunov. Yury, you did a great job by processing the GC-MS profiles and delivering me the ready-to-use data. Moreover, you have been a great help in my volatile understanding, answering all my questions about tentative compound identification, forward match factors, pathways or GeneMaths errors. You always had valuable contributions to the manuscripts.

De mensen bij PPO (WUR glastuinbouw in Bleiswijk) wil ik ook van harte bedanken voor het organiseren van de smaakproeven. Logistiek is het een hele onderneming geweest en jullie wendbaarheid heeft zeker bijgedragen aan het eindresultaat. Specifiek noem ik Monica Kersten, als stabiele factor binnen het 'geweld' van de diverse smaakpanels. Wouter Verkerke wil ik bedanken voor het regelen van het hete-peper panel. Zelf proefde je ook mee en uiteindelijk bleken ze toch best heet hè…

Arnaud en Chris, als co-promotoren hebben jullie een belangrijke rol gespeeld in de totstandkoming van dit promotieonderzoek. Arnaud, jij bent altijd een perfect aanspreekpunt geweest voor praktische, organisatorische en natuurlijk wetenschappelijke onderwerpen. Een antwoord liet nooit lang op zich wachten en vaak zaten we goed op één lijn. Chris, wat ben ik blij dat jij gedurende het project er bij bent gekomen. Ook al valt het in de meeste hoofdstukken niet direct op, maar een groot deel van dit proefschrift bestaat uit statistische analyses. Een aantal zelfs waar ik een paar jaar geleden überhaupt nog nooit van had gehoord. Veel mensen kunnen bevestigen dat wiskunde en statistiek altijd mijn interesse (en kunde) hebben gehad, maar de papers van Breiman over Random Forest zijn toch wel taaie kost hoor! Jouw inbreng is van onschatbare waarde geweest. Voor jullie allebei geldt daarnaast dat jullie ook gewoon hele relaxte, toffe mensen zijn. Ik zou het fantastisch vinden om ook in de toekomst betrokken te zijn bij onderzoeksprojecten met jullie. Aanknopingspunten zijn er volgens mij voldoende.

Richard, ik noemde je al eerder als een van de 'founders' van dit project. Ik neem mijn petje af hoe jij als professor aan het hoofd staat van de leerstoelgroep Plantenveredeling. Naast wetenschapper moet je een hardcore manager zijn om de boel academisch en financieel draaiende te houden. Als promotor van dit proefschrift heb ik je inbreng altijd erg gewaardeerd. Middels jouw helikopterview werden de papers aangescherpt en titels meer out-of-the-box. Jij hebt ook Frank Takken (UvA) als extern begeleider benaderd. Frank, het was vanaf het begin al wel duidelijk dat we elkaar niet heel vaak zouden zien. Uiteraard wil ik jou hier ook bedanken voor

de tijd die jij erin hebt gestoken om vanaf de zijlijn kritisch naar mijn voortgang te kijken en me van waardevolle suggesties te voorzien.

Binnen Rijk Zwaan zijn er zoveel mensen te bedanken, dat ik nu al bang ben dat ik iemand vergeet. Weet dat ik ieders inbreng enorm heb gewaardeerd. Ik heb dit project nooit kunnen afronden door de inbreng van heel veel collega's van afdelingen variërend van Biochemie tot Fytopathologie en Moleculaire biologie tot Marketing. Maar natuurlijk ook de afdeling Kwantitatieve genetica, de mensen binnen Veredeling Aubrika, de gewasmedewerkers, de afdeling Business support, Juridische zaken en de directie. Een aantal mensen wil ik specifiek noemen.

Gerald, Suzanne en Femke, met de afdeling Biochemie hebben jullie een gigantische berg werk verzet. Ik herinner me nog de woorden van Suzanne dat de hoeveelheid metingen het eerste jaar '*toch wel een beetje was onderschat…*'. Ik heb op een hele prettige manier met jullie samengewerkt en heb jullie kritische inbreng bij artikelen en discussies altijd erg gewaardeerd. Dat de Tikunov-methode nu ook een begrip op de afdeling Biochemie is geworden is een mooie bijkomstigheid.

Han en Aat noem ik van de afdeling Moleculaire Biologie onderzoek, natuurlijk zijn er meer mensen op deze afdeling en bij Implementatie geweest die waardevol werk hebben verricht. Han, het moet voor jou een hele uitdaging zijn geweest om elke keer weer aan de, niet of slecht geplande, verzoekjes van mij te voldoen. '*Graag nog even die en die merkers draaien op CA-10 boxjes 52-63 voor eind volgende week'* of '*die boxjes uit 2005 moeten bij jou toch nog ergens in de vriezer liggen? Heb ze namelijk toch nog ff nodig'*. Bedankt voor je flexibiliteit. Aat, bedankt voor de coördinatie van met name het merkerwerk dat buitenshuis is gedraaid. De timing was vaak cruciaal en leidde er toe dat ik je ook wel eens tijdens je vakantie moest bellen.

Evert, jouw afdeling Kwantitatieve Genetica bestond vanaf de oprichting lange tijd uit één persoon. Het was dus wel duidelijk dat ik jou moest hebben voor de lastige vragen. Eigenlijk had ik maar één echt lastige vraag, een goede genetische kaart voor de *C. baccatum* afgeleide populatie, maar daar heb ik je dan ook wel vanaf 2009 tot eind 2012, welteverstaan met tussenpozen, mee bezig gehouden. Telkens was ik toch weer niet tevreden met de kwaliteit van de kaart en regelde ik weer extra merkerinformatie, die nog even een kaartpositie moesten krijgen. Bedankt voor alle nieuwe versies!

Op de Aubrika (aubergine en paprika) afdeling heb ik een geweldige tijd gehad. Jair, ik vond je een hele prettige leidinggevende; super eigenwijs, maar altijd voor reden vatbaar, mits ik mijn argumentatie maar op orde had. Binnen het

promotieonderzoek heb je een hele belangrijke rol gespeeld als sparringpartner. We hebben heel wat discussies gehad over een proefopzet of analysemethode, soms zelfs met de nodige stemverheffing (Antoine vroeg zich zelfs eens af of we ruzie hadden), maar verbazend genoeg (een pietsie eigenwijs ben ik namelijk zelf ook wel) werden we het toch altijd eens. Van jou kreeg ik ook de vrijheid om de vrijdagen thuis te werken, ook al paste dat niet helemaal binnen het bedrijfsbeleid. Van alle andere mensen op de afdeling heb ik ook de nodige support gekregen. Jan van V, Albert, André, Anita, hartelijk dank voor jullie input; onder andere de honderden PCR boxjes die jullie voor me hebben gevuld gedurende dit project zal ik niet snel vergeten (en jullie ook niet ben ik bang…). Corno, jij bedankt voor je flexibiliteit. Door de drukte liet mijn planning nogal eens te wensen over, wat het voor jou niet altijd makkelijk maakte. Jeannette, Jan de B, Bianca, Thijs, Johannes, Antoine, Joep, bedankt voor jullie kritische opmerkingen tijdens de smaakproeven, die ik ad hoc op de afdeling organiseerde. Opmerkingen van Jan '*Wat moet je toch met die rommel'* of Thijs '*Getver dit smaakt naar grond*', hebben me erin gesterkt om vooral door te gaan… En wat heb ik ervan genoten toen ook deze heren bij een van de laatste proefsessies aangaven dat er '*toch wel bijzondere en lekkere smaakjes bij zaten'*. Laure, jij bent vanaf het begin binnen de pre-breeding mijn grote steun en toeverlaat geweest. Wat heb jij een werk verzet en wat was het lachen als we bovenop een elektrocar een honderdtal planten langs reden om die paar planten te vinden met apart smakende vruchten. Merci beaucoup! Björn, de laatste paar jaar hebben we op dezelfde kamer gezeten. Jouw Belgische humor wist alles te relativeren. Als genomic breeder, heb je me ook regelmatig van nieuwe inzichten verschaft. Top. Tamara, als opvolger van mijn job heb je een flinke erfenis gekregen. Naast het smaakproject heb je ook alle andere paprika en aubergine pre-breeding projecten van me over genomen. Bedankt dat je me in de laatste maanden, waarin ik alle projecten heb overgedragen, ook de ruimte hebt gegeven om dit proefschrift te kunnen afronden. Daarnaast vond ik het ook een hele leuke tijd om zo samen te werken. Zsolt, Adriana, Patricia, Celso, Csaba and Arzu, thanks for your continuous support and interest in the project.

Richting de afronding van het project kwam de afdeling marketing meer in beeld. Bauke, Florent en Annette bedankt voor het organiseren van het consumentenonderzoek. Echt een leuke tak van sport en gaaf om te zien hoe mensen reageren op de nieuwe smaakjes. Ook voor jullie nu een uitdaging om dit in de markt te gaan zetten en er een naam aan te hangen… Kevin, Mirjam en Anneke van de afdeling Business Support, bedankt voor het vele werk dat jullie in de staart van het project hebben gestoken. Vanwege de uiteindelijk toch nog strakke deadline van 1 juli, heeft het jullie ook heel wat avonduurtjes gekost. Super, dat het toch allemaal op tijd klaar kwam.

Emilie, merci beaucoup, that you could do the editing of the thesis. It must have been a hell of a job… Tjipke, bedankt voor de gave foto's waarmee de cover is gemaakt.

Naast alle genoemde mensen die inhoudelijk bij het project betrokken zijn geweest wil ik ook een aantal mensen in mijn persoonlijk kring bedanken.

Mijn ouders, broer en zus inclusief hun aanhang wil ik bedanken voor hun begrip in periodes dat mijn hoofd niet echt stond naar 'familie perikelen'. Marjanke, ik vind het cool dat je een van mijn paranimfen wilt zijn. Mijn andere paranimf, Thijs, mag in dit dankwoord ook niet missen. Top, dat jij de vragen van de commissie gaat beantwoorden…als echte paprikafiel verwacht ik echt niets minder. De squashboys bedank ik voor de relaxte maandagavonden waarop de drukte van alledag tijdens het squashen en de aansluitende borrel snel vergeten was. Het was gaaf om te merken dat er ook bij mijn andere vrienden steeds meer interesse kwam in paprika's, veredeling, smaak en natuurlijk zaadloosheid… Onze buurtjes Peter en Charlotte bedank ik voor hun oprechte interesse in de voortgang van het project en de relaxte borrel- of spelavondjes. Dat was een goeie manier om in het weekend toch nog even ergens anders aan te denken. Vooral de laatste twee jaar waren de weekenden namelijk toch wel erg kort. Vrijwel mijn hele proefschrift heb ik thuis geschreven van vrijdagmiddag doorlopend tot zaterdagavond of zondagochtend. Het enige vaste ontspanmoment in het weekend was een uurtje Duck lezen op zondag op de bank. Voor Anita is dit niet altijd even gezellig geweest, ook al heeft ze dit vrijwel nooit laten merken. Anita, naar jou gaat dan ook mijn grootste dank uit. Zonder jouw oneindige steun had ik het nooit gered om dit project af te maken. De vele keren dat we in het weekend samen zaten te eten, waarbij je net zo goed tegen een muur had kunnen praten, omdat ik met mijn hoofd alleen maar bij mijn proefschrift was. De vele uurtjes die je alleen in het weekend doorbracht, omdat ik weer zat te schrijven. Nooit heb je geklaagd omdat ik werkelijk niets in het huishouden deed. Anita, hoe ik het ooit ga goedmaken weet ik nog niet, maar weet wel dat ik oneindig veel van je hou.

Pieter Martijn Eggink was born on December 13th, 1982 in Oostvoorne (The Netherlands). After the completion of his academic high school (Gymnasium) in 2001 at CBSG Blaise Pascal in Spijkenisse, he studied Plant Sciences at Wageningen University (The Netherlands). In January 2007 he obtained his Master degree (Cum Laude) in Plant Sciences with specialization Plant Breeding and Genetic Resources. He performed his MSc thesis at the department of Plant Breeding of Wageningen University, under the supervision of Dr. Pim Lindhout and Dr. Benoit Gorguet. This project focused on parthenocarpy (fruit formation without fertilization/ seeds) in tomato. It was during this project that he realized that a PhD with focus on applied plant research would be an interesting option. In that same period, however, he was also contacted by the vegetable breeding company Rijk Zwaan to apply for the job Pre-breeder *Solanaceae*. In consultation with Rijk Zwaan it was agreed that the position could be combined with a PhD project. In February 2007 he started as Pre-breeder and in January 2008 he started his PhD study connected to the department of Plant Breeding at Wageningen University, under the supervision of Dr. Arnaud Bovy and Dr. Chris Maliepaard on the topic: Taste of pepper. The project was sponsored by his employer Rijk Zwaan. The outcome of this study is presented in this thesis. From July 2013 he is working as Breeding manager tomato at Rijk Zwaan (The Netherlands).

Education Statement of the Graduate School

Experimental Plant Sciences

Issued to: **Martijn Eggink**
Date: **6 November 2013**
Group: **Plant Breeding, Wageningen University & Research Centre**

1) Start-up phase		*date*
‣	**First presentation of your project**	
	Taste in pepper: Characterization of volatile and non-volatile compounds of fresh bell pepper (Capsicum annuum) in relation to sensory evaluation of flavor	Nov 26, 2008
‣	**Writing or rewriting a project proposal**	
	Taste in pepper: Characterization of volatile and non-volatile compounds of fresh bell pepper (Capsicum annuum) in relation to sensory evaluation of flavor	Jul 19, 2008
‣	**MSc courses**	
‣	**Laboratory use of isotopes**	
	Subtotal Start-up Phase	*7.5 credits**
2) Scientific Exposure		*date*
‣	**EPS PhD student days**	
	Intern. PhD Retreat (Joint activity of EPS, SDV & MPIZ), Wageningen University	Oct 02-03, 2008
	EPS PhD student day 2011, Wageningen University	May 20, 2011
‣	**EPS theme symposia**	
	EPS theme 2 symposium 2009, Utrecht University	Jan 22, 2009
	EPS theme 3 symposium 2010, Leiden University	Feb 19, 2010
	EPS theme 2 symposium 2011, University of Amsterdam	Feb 03, 2011
‣	**NWO Lunteren days and other National Platforms**	
	NWO-ALW meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 06-07, 2009
	NWO-ALW meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 19-20, 2010
	NWO-ALW meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 04-05, 2011
‣	**Seminars (series), workshops and symposia**	
	attendance seminars organized by PBR, RZ or Keygene (estimated 20x)	2008-2012
	Research day Plant Breeding, WUR 2008	Jun 17, 2008
	Research days Rijk Zwaan 2008	Nov 26-27, 2008
	Research day Plant Breeding, WUR 2009	Mar 03, 2009
	Research days Rijk Zwaan 2010	Nov 24-25, 2010
	Research days Rijk Zwaan 2011	Nov 07-08, 2011
	Research days Rijk Zwaan 2012	Nov 20-21, 2012
‣	**Seminar plus**	
‣	**International symposia and congresses**	
	XVIth EUCARPIA Tomato work group meeting	May 13-15, 2008
	5th Solanaceae Genome work shop Cologn, Germany	Oct 13-15, 2008
	6th Solanaceae Genome work shop Delhi, India	Nov 09-12, 2009
	XIVth EUCARPIA Meeting on Genetics and Breeding of Capsicum and Eggplant	Aug 30-Sep 01, 2010
	8th Solanaceae Genome work shop Kobe, Japan	Nov 28-Dec 01, 2011
‣	**Presentations**	
	Oral presentations Rijk Zwaan Research Days (4x)	2008-2012
	Oral presentation Eucarpia 2010	Aug 30-Sep 01, 2010
	Oral presentation Intl postgraduate course Plant Breeding	Oct 12, 2012
‣	**IAB interview**	Dec 04, 2009
‣	**Excursions**	
	Subtotal Scientific Exposure	*20.4 credits**
3) In-Depth Studies		*date*
‣	**EPS courses or other PhD courses**	
	Statistical analysis of -omics data	Dec 08-11, 2008
	Metabolomics course	April 12-16 2010
‣	**Journal/discussion club**	
	Member of a discussion group at Rijk Zwaan (Prebreeding)	2008-2013
‣	**Individual research training**	
	Subtotal In-Depth Studies	*4.5 credits**
4) Personal development		*date*
‣	**Skill training courses**	
	NWO talent class; Kernachtig formuleren	Sep 22, 2009
	Project management (NIBI course at RZ)	Dec 08, 21-22, 2010
	Cursus Leidinggeven (RZ course)	May 24 & Jun 01, 2011
	Leiderschapsprogramma GROWTH (Schouten en Nelissen)	Apr-Jun 2013
‣	**Organisation of PhD students day, course or conference**	
‣	**Membership of Board, Committee or PhD council**	
	Subtotal Personal Development	*3.7 credits**
	TOTAL NUMBER OF CREDIT POINTS*	**36.1**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

** A credit represents a normative study load of 28 hours of study.*

This research was financed by Rijk Zwaan.

Cover:

Design: Emilie Fradin
Pictures: Tjipke Meijer, Anita Eggink

Printed by Wöhrmann Print Service, Zutphen, The Netherlands.