INTROGRESSION OF GENETIC RESISTANCE TO DOWNY MILDEW (*PERONOSPORA BELBAHRII*) IN A NON-MODEL PLANT SPECIES, SWEET BASIL (*OCIMUM BASILICUM*)

By

ROBERT M. PYNE

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Plant Biology

Written under the direction of

Dr. James E. Simon

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2017

ABSTRACT OF THE DISSERTATION

Introgression of Genetic Resistance to Downy Mildew (*Peronospora belbahrii*) in a Non-Model Plant Species, Sweet Basil (*Ocimum basilicum*)

By ROBERT M. PYNE

DISSERTATION DIRECTOR:

James E. Simon

Sweet basil (*Ocimum basilicum*) is among the most widely cultivated culinary herbs in the United States, Western Europe and Israel. Despite relative economic importance, breeding and genetic study of this plant species has been largely neglected, rendering its >3 Gbp genome largely unexplored. The deficit in available *O. basilicum* genetic and genomic resources has been highlighted by a worldwide downy mildew epidemic caused by obligate oomycete *Peronospora belbahrii*, which has yet to be mediated by disease resistant varieties. The goal of this dissertation research was to narrow the information gap preventing an effective disease resistance breeding response. Specific objectives were to: 1) determine population structure and estimate genetic diversity among a panel of downy mildew resistant and susceptible *Ocimum* spp. accessions, 2) identify mode of inheritance for resistance to downy mildew in a full-sibling family and 3) construct a linkage map for detection of quantitative trait loci (QTL) associated with DM resistance. A nested, model-based cluster analysis demonstrated three major delineations within the *Ocimum* genus with additional evidence for cryptic structure, especially within the economically important k1 *O. basilicum* cluster.

Distribution of DM resistance was concentrated outside the k1 *O. basilicum* cluster with the exception of a single k1 genotype, 'MRI'. Analysis of downy mildew response across F_2 and backcross populations over two years and two locations demonstrated major gene control of downy mildew resistance conferred by MRI. Finally, a restriction site associated DNA sequencing (RADseq) approach facilitated the discovery and mapping of >1,800 single nucleotide polymorphism (SNP) and expressed sequence tag simple sequence repeat (EST-SSR) markers. The resulting genetic map was validated by the detection of a major QTL, *dm11.1*, which explained 38-55% of the phenotypic variance observed for the MRI x SB22 F_2 mapping population. Disomic inheritance of SNP and SSR markers support previous cytological evidence that basil has evolved an allopolyploid genome. Results of this dissertation provide the most robust phylogenetic examination of the *Ocimum* genus to date, characterization of DM heritability across multiple environments and the first report of genetic/QTL mapping for *O. basilicum*. A current case study is provided for the feasibility of breeding a non-model plant species using classical genetic theory in combination with modern genomic technologies.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Jim Simon my advisor, mentor and friend. Dr. Simon's consistent encouragement and confidence in my ability to navigate the scientific process was essential to the execution of this dissertation. I would also like to thank my committee members Dr. Andy Wyenandt, Dr. Stacy Bonos and Dr. Todd Wehner who were excellent sources of expertise in plant genetics, breeding, pathology and general biology. Access to Dr. Bonos and Dr. Honig's laboratory, which included assistance from Jennifer Vaiciunas and Chrissy Kubik, was essential to the success of this research. Dr. Adolfina Koroch and Dr. Rodolfo Juliani provided intellectual and moral support essential from the genesis to the completion of this dissertation. In addition, I am fortunate to have had the opportunity to learn and seek counsel from numerous scientists including Dr. Richard Raid, Ann Hartman, Dr. Meg McGrath, Dr. Rob Wick, Dr. Li-Jun Ma, Dr. Bingru Huang, Dr. Dan Ward and Dr. Rong Di.

I thank the Rutgers Agricultural Experiment Station, which facilitated a great deal of my research. In particular I thank Jack Rabin, Ed Dager, John Grande and the entire Snyder Farm staff who were instrumental in sparking my interest in agricultural research. I would like to thank once again Dr. Andy Wyenandt and the Rutgers Agricultural Research and Extension Center (RAREC) staff for assistance in field experiments. I am grateful to Joe Florentine, Mark Peacos and research greenhouse staff for their horticultural expertise and assistance in maintaining plant germplasm.

It has been a privilege to work in the New Use Agriculture and Natural Products Program (NUANPP). As such I owe gratitude to the NUANPP laboratory as well as all members with whom I have been lucky to work. In addition I would like to thank my

fellow graduate students Dr. Megan Muehlbauer, Dr. Lisa Beirn, Dr. Greg Behringer, Vincenzo Averello, Dan Giurleo, William Reichart, Dave Byrne, Kathryn Homa and many others. I thank the various undergraduate students and technicians who assisted in my research including Tatiana Quiroz, Baylee Carr, Kathryn Kingsley, Lisa Miamone and Carly Seldow.

I am fortunate and grateful to have received financial support in the form of assistantships from USDA special crops research initiative (SCRI) and the Binational Agricultural Research and Development (BARD). In addition, I would like to thank the Department of Plant Biology for support in the form of scholarships that assisted in the execution and presentation of findings at national meetings.

I am especially grateful to my family and friends. I would like to acknowledge my mother and father, Helen and Niall, whose sacrifice and hard work afforded me a life in which to pursue what I love. Collective journeys that began in rural Ireland, but spanned from midwifery in Saudi Arabia to a label printer factory outside Cologne to apartheid South Africa all in pursuit of a fulfilling life remain a constant inspiration in and outside my career. I dedicate this dissertation to my parents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1.

Literature Review

Sweet Basil: A Unique Specialty Crop

Sweet basil (*Ocimum basilicum* L. Lamiaceae) is an annual herbaceous plant species with a specific epithet derived from the Greek word vasilikós or 'royal' according to the Oxford English Dictionary. This title maintains contemporary relevance as sweet basil continues to demonstrate worldwide demand as a high-value specialty crop used in the horticultural, culinary, perfumery and food flavoring industries [1]. Primary production locations include the southeastern and western United States [2], Mediterranean Europe, Israel, North Africa, India, South-East Asia [3]. US production area and market value statistics are complicated by the inclusion of sweet basil in the multi-crop class of 'culinary herb' by the National Agricultural Statistics Service [4]. Consolidation into this class prevents direct production estimates, however, sweet basil contributes to the annual 9,045 acres of fresh market culinary herbs produced and $96,796 sold as bedding plants according to the 2012 U.S. Department of Agriculture Census (https://agcensus.usda.gov). Italy is considered the top European consumer of sweet basil [3] and produces 4,500-4,900 tons/year [5], while in Israel sweet basil is the primary herb crop with an estimated 11,000 tons/year [6].

A distinguishing feature of basil (*Ocimum* spp.) is its robust biosynthetic network responsible for a diverse array of secondary metabolites, many of which confer human health benefits. For instance, radical-scavenging polyphenols accumulate at high density in sweet basil leaves, providing a rich source of antioxidant activity [7]. The predominant phenolic, rosmarinic acid [7,8], has been shown to provide anti-inflammatory activity and

significantly decrease viral loads in mice [9]. A second important class of sweet basil secondary metabolites is the volatile organic compounds (VOCs), which are produced in essential oils and comprised of two distinct chemical groups: (i) terpenoids (mono- and sesqui-) and (ii) phenylpropenes. These low-molecular weight compounds are considered to be the result of evolutionary fitness as a plant-pathogen defense system [10]. Unsurprisingly, essential oil extracts have been shown to demonstrate antiviral [11], antibacterial [12] and antifungal [13] properties. Volatile compounds in sweet basil are of greatest commercial importance as the major components responsible for flavor and aroma.

The relative abundance of terpenoid and phenylpropanoid volatiles found in sweet basil essential oils collectively constitutes a chemotype. A multitude of sweet basil (*O. basilicum*) essential oil compositions have been characterized [1,14–16] including those rich in monoterpenoids (linalool, 1,8 cineole, geraniol, geranial, neral) and phenylpropenes (methyl chavicol, eugenol, methyl eugenol, methyl cinnamate). Demand for specific chemotypes is largely dependent upon geographic region [1], however, the predominant and most economically important is the linalool-rich Western European or 'traditional' sweet basil type. Although organoleptic, social and demographic differences among human populations prevent a consensus or optimal volatile composition for the 'traditional' chemotype, the most widely produced cultivars are dominated by linalool (50-70%) and complemented by the minor components methyl chavicol (≤30%), 1,8 cineole (≤13%) and eugenol (≤13%) [17].

Sweet basil volatile chemistry is a unique trait for which there is no substitute in any other plant species. Greatest economic demand for sweet basil occurs in the fresh and

culinary markets where it is utilized primarily for its unique specific flavor profile. Therefore, abiotic or biotic factors that threaten production of sweet basil endanger access to a food product for which there is no replacement.

Basil Phylogeny

Phylogenetic classification within the basil genus is complicated by natural and artificial outcrossing [18] resulting in extensive variation in genetics, morphology, secondary metabolite chemistry, and ploidy [19]. Taxonomic analysis by Paton et al. [20] recognized 64 *Ocimum* spp. using various morphological descriptors to partition the genus into three subgenera: *Ocimum*, *Nautochilus* and *Gymnocinum*. Sweet basil (*O. basilicum*) belongs to section *Ocimum* within subgenus *Ocimum*.

More recently, additional traits have been incorporated into systematic evaluation of the genus including aromatic volatiles, flavonoids, and molecular markers [14–16,21]. Studies demonstrate that individual chemotypes can be represented by multiple plant morphologies, which has resulted in confusion and mislabeling of cultivars especially within the *O. basilicum* species [20]. A number of genetic diversity studies have been conducted using plastids [22] random amplified polymorphic DNA [23], and amplified fragment length polymorphism DNA markers [24] in conjunction with morphological and volatile composition data to better define phylogenetic relationships among and within species. However, a more extensive characterization of the relationship among basil ac- cessions and/or varieties as well as greater clarification as to origin and species is greatly needed. In addition to differences in genotype, *Ocimum* spp. exhibit variation in genome structure. Cytological and nuclear DNA content studies provide evidence for different basic chromosome numbers among basil species [25–27]. Sweet basil is

considered to be a tetraploid with 2n = 4x = 48 [27], while *O. americanum* and *O. africanum* are most likely hexaploids (2n = 6x = 72) [27,28]. Flow cytometric analysis of 23 basil genotypes representing eight species found a wide distribution of genome sizes ranging from 0.92 Gbp (*O. campechianum*) to 5.5 Gbp (*O. americanum*) with sweet basil (*O. basilicum*) being within the 2.97 to 3.39 Gbp range [29]. Wide differences in genome constitution among species result in barriers to sexual reproduction, which has limited interspecific crossing [30] and obstructed downy mildew resistance gene introgression (J.E.Simon, R.M.Pyne, and C. A.Wyenandt, unpublished data).

The Downy Mildew Pathogen with a Focus on *Peronospora belbahrii*

Few plant-pathogenic diseases rival the economic losses due to crop destruction as the downy mildews (DM). An estimated 16.7% of the 6.25 billion USD global fungicide market is dedicated to DM control [31]. DM belongs to the oomycetes, a diverse class of straminipilous organisms originally believed to be true fungi but later shown to be more closely related to diatoms and brown alga of the phylum Heterokonta [32]. Oomycetes evolved in parallel to true fungi and have a number of convergent morphologies but are distinguished by multiple characters, most notably in hyphal walls composed primarily of cellulose and β–(1,3)-glucans [33]. Multiple notorious diseases are caused by a variety of pathogenic oomycete genera (*Phytophthora* sp., *Pythium* sp. and *Albugo* sp.) and more than one third of oomycete species belong to the 19 genera described for the DMs (Order: Peronosporales; Family: Peronosporaceae) [34]. The DMs constitute a large group of obligate biotrophs that include >700 species in a monophyletic grouping nested within the *Phytophthora* genus. However, further phylogenetic

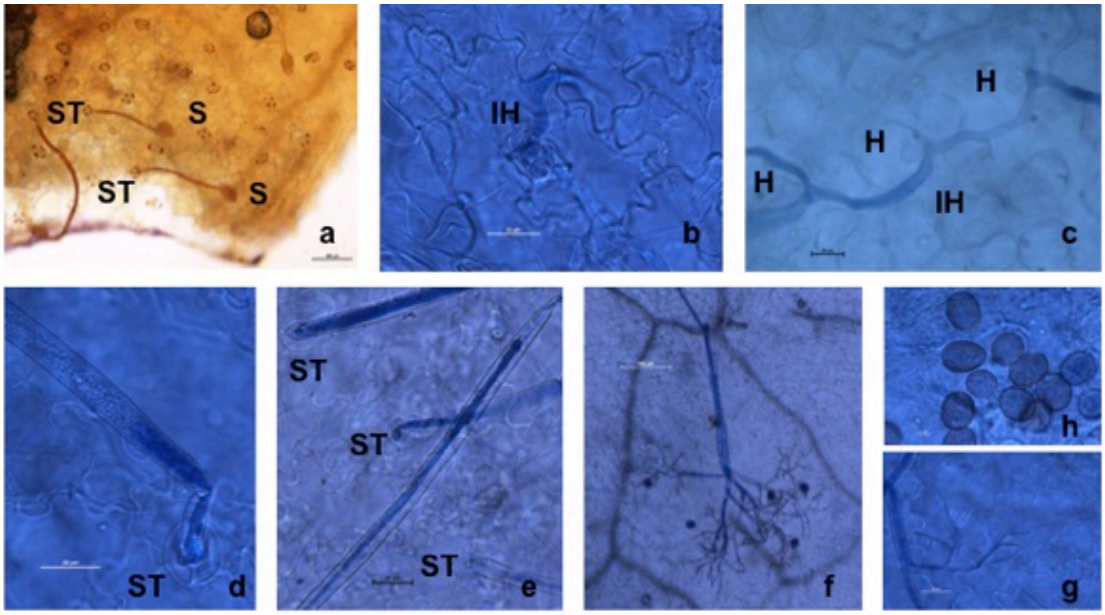
investigation of a more representative sampling of the DM genera is needed to better characterize their relationships [34].

The DMs can be broadly divided into three major groups classified according to 1) colored conidia, 2) pyriform haustoria and 3) Brassicaceae host range [34]. The first and third groups contain the model pathogens *Pseudoperonospora cubensis* (cucurbit downy mildew) and *Hyaloperonospora arabidopsis* (a model for *H. parasitica* of crucifers), respectively, while the second group contains the economically salient lettuce *Bremia lactuca* (lettuce), *Plasmopara viticola* (grape) and *Plasmopara halstedii* (sunflower) DMs [34]. *Peronospora* sp. (group one) is the largest DM genus with a species defined as a function of host range. Conflicting opinions of phylogenetic classification persist with regard to whether broad [35,36] or narrow host-specificity should define *Peronospora* species delineations. Molecular evidence points to a prevailing trend of *Peronospora* sp. being defined according to the latter philosophy. For instance, analysis of ITS rDNA sequences and morphologies among spinach DM isolates by Choi et al. [37] provided evidence of a new species-specific (*Spinacia* spp.) DM pathogen *P. effusa* as opposed to the previously described *P. farinosa*, having a host range that included the entire Chenopodiaceae family. However, *Pseudoperonospora cubensis*, believed to infect 20 genera of Cucurbitaceae [38], provides a more broad species definition. Classification of the genus is complicated by rapid evolution [39] and frequent pathogen 'host-jumping' between non-related host families such as *P. cubensis* pathotype 9, which infects hop (*Humulus lupulus*) in addition to *Cucumis* sp. These complexities have prevent *Peronospora* sp. from being determined monophyletic [40].

Basil downy mildew (BDM) is caused by *Peronospora belbahrii* [41] and has a genus-specific host-range restricted to *Ocimum* spp. [42]. The earliest report of BDM was in 1933 in Uganda [43] and was not documented again until 2001 in Switzerland [41] then eight years later in North America [44]. Since then *P. belbahrii* has spread rapidly throughout the world [45–48] becoming a worldwide disease epidemic most likely associated with pathogen transmission through infested seed [49]. Contaminated seed is a major disease management issue for other DM pathogens including *P. farinosa* f. sp. *spinaciae* (syn. *P. effusa*) [50] and *P. cubensis* [51]. *P. belbahrii* has also been detected with endpoint PCR in live, asymptomatic plants [52] and is therefore likely to be disseminated by infected seedlings. Versatility of the BDM pathogen, despite its obligate nature, makes it a devastating disease with enormous economic impact. A precise understanding of the *P. belbahrii* life cycle and the required environmental conditions is thus critical to implementation of control measures.

The life cycle of *P. belbahrii* and all other DM pathogens is initiated as a sporangium, an asexual spore that is readily dispersed by water or wind. In contrast to the basil lineage of DMs in which sporangia release motile zoospores [33,34], sporangia of *P. belbahrii*, as well as *Peronospora* and *Pseudoperonospora* sp., produce a germ tube upon contacting the susceptible host leaf surface (Fig. 1a,g). The germ tube will proceed to locate and penetrate the stomata, gaining access to the leaf mesophyll (Fig. 2a). Sporangia germination and location of stomata in an aqueous environment were shown to take place within 6 and 24 h, respectively, in the absence of light, at 15°C [53]. Once inside the leaf mesophyll, intercellular mycelia proliferate, surrounding the host plant cells and producing globuse haustoria (a hallmark oomycete structure) for nutrient acquisition (Fig

1c.). The haustorium will invaginate the host cell, forming an extra-haustorial matrix into which effector proteins will be secreted to suppress the host defense response and promote infection [54]. Finally, a mycelial cushion is formed in the substomatal cavity [33] that provides the physical foundation for the production of sporangiophores or sporangia-bearing fruiting bodies that emerge from stomata (Fig 1d-f). Sporulation (i.e.,



Peronospora belbahrii morphology and stages of pathogenesis. (a) Sporangia directly germinate and enter host through host stomata (b) Hyphae penetrate through the lower epidermis .(c) Hyphae advance intercellularly through mesophyll tissue. Haustoria invaginate mesophyll cells for nutrient acquisition. (d,e) Sporangiophores exit through stomata and produce secondary inoculum (sporangia) from the distal end of the hyphal tips. (f) Sporangia. (g) Dichotomous hyphal branching of sporangiophore. H, Haustoria; IH, Intercellular hyphae. S, Sporangia. ST, stomata.

sporangiophores) is a critical life cycle stage for DMs as it marks the commencement of the polycyclic disease cycle, resulting in exponential inoculum proliferation and ultimately disease epidemic given favorable conditions [2,55].

Environmental conditions required for sporulation vary among DM species, but generally require sustained periods of leaf wetness, high humidity and moderate temperatures [33,56,57]. For BDM, the optimal temperature for *P. belbahrii* sporulation is 20°C [53,58,59], but will occur at 10-25°C to varying extent depending on leaf wetness and relative humidity. Relative humidity of >85.1% [53] and leaf wetness (dew period) of

at least 24 hours after symptom development [59] are required for sporulation. In controlled greenhouse experiments, sporulation occurs 6 – 7 [53,58,60] days post inoculation (dpi).

The sexual phase for DM generally takes place in leaf tissue and requires compatible mating types to form gametangia [33]. While *P. belbahrii* mating types have not been observed, study of model systems *P. viticola* and *P. cubensis* suggest DM mycelia are primarily heterothallic requiring the presence of A1 and A2 mating types for sexual reproduction [33,61]. Sexual reproduction is an essential tool used by DM pathogens to facilitate survival (i.e., overwintering) when subjected to increased selection pressure.

Intensive use of mefenoxam for control of DM in Europe has resulted in recent reports of resistance of *P. belbahrii* isolates in Israel [62] and Italy [63,64]. Cohen et al. [16] identified oospores in leaves of susceptible cultivar 'Peri' which had been treated with 1,000 μg Mefenoxam/ml yet exhibited sporulation in field and greenhouse experiments. Metaxyl-based control of *Bremia lactucae* became ineffective in the 1980s as a function of evolved pathogen insensitivity [65] and will inevitably become obsolete for basil DM management. Furthermore, over time *P. belbahrii* population diversity is expected to diversify greatly with increased external selection pressures and among geographic locations as has been evidenced by the enormous genetic diversity exhibited by the cucurbit DM pathogen [55].

Plant-Pathogen Interaction and Host Resistance

Plant disease resistance can be most broadly classified as either nonhost or host resistance [66]. The former case is most ubiquitous as it is characterized simply by the

more likely occurrence of an incompatible relationship between nonhost species and a given pathogen resulting in plant immunity [33]. Mysore and Ryu 2004 define type I nonhost resistance as asymptomatic and due to the obstruction of a pathogen from entering the host cell. This may be a function of nonhost morphological features (e.g., leaf cuticle, phytoalexins, secondary metabolites, etc.) serving as physical obstacles or as a generic elicited response to foreign molecules such has been demonstrated by the conserved bacterial flagellin protein domain flg22 [67]. In the latter scenario, an instance of pathogen-associated molecular patterns (PAMP) identified by transmembrane plant recognition receptors (PRRs) results in PAMP-triggered immunity (PTI) or phase 1 of the zigzag model for plant immunity [68].

Type II nonhost resistance occurs when the pathogen surpasses the type I basal response or PTI and invades the host cell by secretion of effector proteins. Effector-triggered susceptibility (ETS) (phase 2) takes place in the absence of host resistance (R) genes that encode proteins for recognition of such effectors [68]. In contrast, an effector-triggered immunity (ETI) (phase 3) is produced in a nonhost with R genes encoding cytoplasmic nucleotide binding site-leucine rich repeat (NBS-LLR) domain proteins. This circumstance can be phenotypically distinguished from type I nonhost resistance by occurrence of necrosis through programmed cell death or hypersensitive response (HR) [69].

Host resistance and type II nonhost resistance responses are phenotypically indistinguishable, both characterized by HR as well as localized production of reactive oxygen species (ROS) [66]. In a host resistance response, the pathogen secretes effector proteins, deemed avirulence (Avr) proteins when recognized by host NBS-LRR proteins

during ETI. This compatible interaction is better known as the ‘gene-for-gene’ concept [70,71]. The line between host (gene for gene) resistance and nonhost resistance is often blurred from a molecular standpoint as substantial commonality is found in the genetic mechanisms required for both responses [66]. Instances are documented in *A. thaliana* where the *ENHANCED DISEASE SUSCEPTIBILITY 1* (*EDS1*) gene was first identified as a nonspecific locus activating basal defense (PTI) to *Hyalaperonospora parasitica* [72]. The *PHYTOALEXIN DEFICIENT 4* (*PAD4*) was later shown to interact with *EDS1* to both promote salicylic acid (SA) accumulation in a positive feedback loop facilitating PTI [73]. In the event that *H. parasitica* is able to subvert this basil defense response, *EDS1* and *EDS1-PAD4* have also been shown to recruit an N terminal Toll–interleukin-1–receptor domain (TIR) type NBS-LRR protein for ETI [74]. A similar type II nonhost response was observed following invasion of *Arabidopsis* host cells by *Pseudomonas syringae* effector AvrRps4 where *EDS1* was shown to act as a target, binding AVrRps4 for activation of TIR-NBS-LRR resistance protein RPS4 in a post-invasion defense [75]. Redundancy of host molecular machinery demonstrates the complexity of disease response and a consideration is essential to resistance development strategies [66]. Nevertheless, host resistance is typically associated with genes, denoted *R* genes, which provide defense against specific pathogen races. This form of defense is particularly common in the host-DM pathosystem [66].

The intimate interaction between DM pathogens and their host is a function of obligate biotrophy, a defining characteristic that has accelerated pathogen-host evolution [34]. This relationship is physically manifest through formation of a unique extrahuastorial membrane that replaces the host cell plasma membrane [76]. Haustoria

secretion of an ever-diversifying array of effector proteins [77] has placed intense selective pressure on the host to produce compatible NBS-LRR proteins [78]. This mutual arms race has resulted in plant genomes rich in NBS-LRR *R* proteins including 149 in *Arabidopsis thaliana*, 228 in *Manihot esculenta* [79] and as many as 1015 in *Malus x domestica* [80].

The most well studied group of DM *R* genes is the resistance to *Hyaloperonospora parasitica* (RPP) class of Arabidopsis. Six RPP genes (RPP5, RPP8, RPP1, RPP7, RPP13 and RPP2) have been cloned [81–86] belonging to either the TIR or coiled coil (CC) N-terminal domain NBS-LRR protein sub-families and all of which trigger ETI and result in HR. Combined quantitative trait loci (QTL) analysis and genome wide association study (GWAS) demonstrate *Hyalaperonospora arabidopsis* resistance is controlled by single dominant gene action conferred by loci that maps to four genomic regions [87]. Similar *R* genes have been identified across plant families, including the CC-NBS-LRR encoding *Dm3* gene cloned in *Lactuca sativa* cv. Diana and shown to be part of a larger *RGC2* gene family [88]. Wild North American grapevine species *Muscadinia rotundifolia* TIR-NBS-LRR encoding *MrRPV1* was shown to significantly reduce *Plasmopora viticola* sporangia production following transient expression in *V. viticola* [89]. In sunflower, RBC151 and RGC203 genes were found to encode TIR- and CC-NBS-LRR proteins involved in two unique mechanisms of *Plasmopara halstedii* resistance [90].

Breeding for Qualitative Downy Mildew Resistance

The ubiquity of single dominant NBS-LRR encoding genes have been exploited knowingly and prior to cloning efforts through classic genetics and breeding approaches.

In spinach a single resistance locus *Pfs* is deployed in resistant germplasm as a primary means of disease management against *P. farinosa* races 1 and 2 [91]. Similarly, the *Cucumis melo Pc-3* [92] and *L. sativa Dm3* [88] loci have been utilized in DM resistance breeding programs for decades [93,94]. DM resistance is also commonly conferred by recessive gene such as the 2OG-Fe(II) oxygenase encoding *Arabidopsis* DMR6 gene demonstrated to increase resistance in recessive mutants [95]. Cucumber breeders in the 1950s began incorporating HR-associated DM resistance from USDA-GRIN plant introduction (PI) 197087 [96], which was later characterized as a single recessive locus *dm-1* [93].

Introgression of host specific, single dominant or recessive loci into cultivars is an appealing strategy for rapid introduction of strong resistance to DM, however, it is prone to rapid breakdown from minor mutations such as in *Avr* effector encoding genes resulting in new pathogen races [97]. Cucurbit breeders now contend with six major pathogenic races of *P. cubensis* in the United States including a new and devastating *dm-1* resistant strain that emerged in 2004 [38,98]. In the absence of known host resistance to this new strain, disease epidemic and soaring fungicide use (ie. chlorothalonil-treated acreage doubled between 2006 and 2010) continually plagued cucumber production until the 2012 release of improved resistant cultivars [98]. Disease resistance breeding has become increasingly sophisticated, mediating such events through incorporation of multiple sources of resistance for more durable resistance [99]. Parra et al. [94] distilled the publically available lettuce DM resistant gene bank into 51 single Mendelian loci and *R*-factors (not yet shown to be monogenic or mapped) to provide a non-redundant catalog of resistance sources for combined deployment in breeding. Evaluation of F_2 and

backcross (BC_1) generations derived from doubled haploid *B. oleraceae* accessions with differential response to 42 *H. parasitica* isolates crossed with DM susceptible cultivars revealed multiple single, dominant genes among families [100]. Combination of single gene, race-specific resistance (gene pyramiding) is an advantageous breeding strategy because it broadens the compatible gene-for-gene response across a wide range of pathogen diversity often associated with multiple locations [55].

Breeding for Quantitative Downy Mildew Resistance

Sources of DM resistance may also be inherited as a polygenic or quantitative trait in which the level of resistance conferred is dependent upon the allelic state of multiple loci in the host genome. In this scenario, the gain from selection of favorable alleles can be simply expressed in what is termed the 'breeder's equation' or $\Delta G = ir\sigma_a$, where I = selection intensity, r = accuracy of the selection and σ_a is the standard deviation due to additive effects across individuals [101,102]. For qualitative disease resistance ΔG is simplified to 100% (presence of resistance allele(s)) or 0% (absence of resistance allele(s)). Thus, qualitative disease resistance breeding typically fixes *R* gene alleles in early generations to facilitate subsequent selection for polygenic traits among resistant germplasm [102]. Quantitative resistance demonstrates variation across phenotyping experiments with fluctuating values of σ_a [103]. Hence, the phenotypic variance explained (PVE) across locations, years, etc. are often different, which reduces ΔG. However, this form of resistance is not confined to the 'gene-for-gene' paradigm and therefore not dependent upon recognition of a single pathogen effector as in a host resistance response.

DM resistance in *Humulus lupulus* is currently attributed to quantitative resistance as described by Henning et al. [104] in which narrow sense heritability (h^2) of resistance conferred by the cultivar 'Teamaker' ranging from 0.38 to 0.57 across environments. This study is indicative of the difficulties inherent in development of a reliable gain from selection model for polygenic DM resistance across locations. Twenty-two single nucleotide polymorphisms (SNPs) were linked to a DM resistance response, or quantitative trait loci (QTL) and identified with varying PVE [104]. In this circumstance, a most feasible approach may be to focus breeding efforts on QTL with the highest LOD scores as suggested in by Henning et al. [104] for three QTL with logarithm of odds scores (LOD) scores > 12.03. Investigation of quantitative resistance in cucumber to the closely related *P. cubensis* pathogen demonstrates greater stability. F_2 and BC_1 generations from a cross between DM tolerant Ames 2354 (derived from self-pollination of PI 234517) and DM susceptible PI 175695 exhibited relatively high heritability in multiple greenhouse evaluations using a Polish *P. cubensis* isolate, providing evidence for an approximate 20% gain from selection [105]. Furthermore, a two-year study across environments in three countries detected five QTL that explained 62-76% of phenotypic variance inherited from post-2004 *P. cubensis*-strain resistant accession Q17120 (derived from 330628) [106]. Given the diversity of environments, these results suggest promise for cucumber polygenic DM resistance breeding.

Substantial progress has been achieved in breeding for polygenic resistance in lettuce where 15 QTL have been identified to date [94]. Lettuce (*L. sativa*) legacy cultivars 'Iceberg' and 'Grand Rapids' were both developed over a century ago and demonstrate quantitative resistance with high h^2 [107]. A recombinant inbred line (RIL)

population derived from a cross between these two cultivars was used develop a high density map for QTL discovery [108]. In this study, two QTL were identified on linkage groups 2 and 5 across locations in California and the Netherlands with resistant alleles (*ii*) originating from Iceberg [108]. The most common source of quantitative DM resistance in lettuce is from *L. saligna*, a species that demonstrates *B. lactucae* nonhost resistance [94]. Analysis of an F_2 generation derived from grandparents *L. sativa* (DM susceptible) x *L. saligna* (DM resistant) revealed three significant QTL conferring differential PVE to two separate *B. lactucae* isolates [109].

A recent trend in lettuce breeding is a backcross strategy for development of *L. sativa* near isogenic lines (NILs) with introgressed *L. saligna*, *L. serriola* or *L. virosa* donor genome segments or QTL conferring DM resistance [94]. An initial effort to stack 8 QTLs by intercrossing NILs developed from multiple *L. saligna* resistance sources proved less predictable than expected in a simple additive x additive model (ie. epistasis) [110]. Epistasis is a gene effect that becomes more likely with QTL stacking or pyramiding as addition of each gene corresponds to the exponential addition of gene pairs and thus opportunity for unpredictable interaction. Although this reduced predictable gain from selection, in some cases it may provide opportunity for transgressive phenotypes such as a 'more-than-additive' resistance, which was observed for two NIL combinations [110]. Modern breeding approaches are moving towards multiple disease resistance in plants due in large part to increasingly robust genotyping methods.

Genetics, Genomics and Breeding in a Non-Model System

Plant breeding experienced an enormous leap in the early 1900s following the birth of the genetics and the incorporation of Mendelian heredity into selection strategies

[103]. Early genetic improvement efforts targeted a handful of socio-economically significant plant species such as maize, soybean, sorghum and wheat, primarily for increased yield [103]. Accumulation of genetic resources and characterization of plant species throughout the 20th century was therefore correlated with commercial salience. Plant germplasm, breeding lines (inbred, dihaploid, etc.) and (NILs, RILs, etc.) segregating families for these species have been intensely developed as a function of disproportionately higher resources allocated over a substantially longer period of time. This is perhaps best reflected in the currently available USDA Maize Genetics Cooperation Stock Center (MGCPS) 25 Nested Association Mapping (NAM) populations 5,000 RILs among 25 populations http://maizecoop.cropsci.uiuc.edu/nam-rils.php.

Approximately one century after the discovery of discretely inherited units (genes) and decades after the elucidation of their structure, the 'genomic era' once again transformed plant breeding [111]. In this new era focus was shifted to plant systems with few essential characteristics including a small (<500 Mbp) /diploid genome, short generation time and ease of transformation. Thus, rather than selecting plant species as a function of economic importance, 'model' species related to economically important plant families were selected in the hope that important genes and gene action will be shared with cultivated species. The most prominent example being the Brassicaceae model species *Arabidopsis thaliana*, notably the first plant genome to be fully sequenced (The Arabidopsis Genome Initiative, 2000). Others include the 430 Mbp *Oryza sativa* for modeling cereal crops and the ~500 Mbp *Medicago truncatala* for modeling legumes and the 240 Mbp *Fragaria vesca* [113] for modeling octoploid *Frageria x anassa* and other Rosaceae species.

The 21st century marked a paradigm shift in plant breeding with a new emphasis on DNA genotyping burgeoned by the availability of Next-Generation-Sequencing (NGS). Sequencing costs following the development of NGS dropped precipitously from $397.09/Mbp pair in October 2007 to $3.81/Mbp one year later and $0.014 as of October 2015 (https://www.genome.gov). In parallel to decreased cost per nucleotide, DNA sequencing quality (Q) and read lengths continue to increase. New long-read sequencing technologies continue to emerge such as the PacBio SMRT platform, which is capable of generating an average read length of >10 Mbp (http://www.pacb.com/smrt-science/smrt-sequencing/read-lengths/). Model and economically important species once again have been the initial benefactors of this technology, a point highlighted by the now 1,135 available *A. thaliana* sequenced genomes [114]. Although large-scale WGS for inter-individual variation is not feasible for the majority of plant species, the trickle-down effects of NGS to non-model species are beginning to be realized [115]. Kang et al. [111] reported approximately 99 plant species with draft genome assembled, while the NCBI genome database lists 199 'land plants' with varying levels of assembly (https://www.ncbi.nlm.nih.gov/genome/; accessed January 20, 2017). The explosion of genomic sequencing that has accompanied plummeting costs has greatly diminished the issue of marker throughput, allowing for discovery of SNP DNA markers in the tens of thousands with minimal cost or labor [111].

A number of NGS-based innovations have been developed for incredibly high-throughput genotyping at the population level [116]. Multiple SNP arrays have been developed for automated, multiplexed PCR genotyping using a robust set of DNA markers. The MaizeSNP50 DNA analysis kit features more than 50,000 validated SNP

markers from the B73 Maize reference genome generating an average of more than 25 markers per Mbp (http://www.illumina.com/products/by-type/microarray-kits/maize-snp50.html). In the non-model cultivated species of Rosaceae, an extensive collaboration among multiple research groups (RosBREED) has yielded powerful SNP arrays in apple, peach, cherry and most recently in octoploid strawberry with publication of the 90 K Axiom SNP array generating 36 million variants across 19 accessions [117]. This collaborative effort may represent a model for the development of genomic-assisted breeding (GAB) tools in non-model species.

Although the annual number of published genomes experienced a six-fold increase from 2010 to 2014, the vast majority were for organisms with less than 1 Gbp [111]. Thus, nearly all non-model species with large, often polyploidy, genomes remain without a reference genome due mostly to a bottleneck in bioinformatics (primarily with regard to assembly) [115]. Instead low depth WGS can be used as a quick and dirty approach to mine for SSRs [118], single copy orthologs (COSII) and rDNA loci without a reference genome (ie *de novo*) [119]. PCR-based markers such as SSRs are amenable to multiplexed, population-level genotyping with capillary electrophoresis. Microsatellite markers are particularly suited to genetic diversity and population structure studies, while SNPs have become the marker type of choice for linkage mapping due to their ubiquity throughout the genome [116]. In non-model species the concept of "genome skimming" and massively parallel sequencing for SNP discovery has revolutionized genotyping [115,120,121]. A number of approaches have been exploited including hybridization to probes (ie SNP arrays) [117], 5'-anchored PCR-enrichment [118], exome sequencing

[122], poly(A)+ mRNA sequencing (ie RNAseq) [115] and restriction-enzyme-based enrichment [123].

Among the available genotyping methodologies, restriction site associated DNA sequencing (RADseq)[124], alternatively referred to as genotyping by sequencing (GBS) [121], has become particularly favored in the plant breeding community [125–127]. In RADseq a subset of homologous sequences are sampled from the genomes of a plant population of interest by: 1) digesting genomic DNA, 2) ligating sequencing adaptors/PCR primers/4-9 bp barcode sequences/common Y- adaptors, 3) PCR amplification of fragments with barcode + common Y- adaptors 4) pooling (multiplexing) libraries, 5) sequencing [123]. Ultimately this process provides a reduced genome representation for mining sequence variants (SNPs) across individuals [123]. The original single-digestion protocol with *ApeKI* [121,124] was later modified to the *MspI-PstI* double-digestion (ddRADseq) to map 34,000 and 20,000 SNPs in the barley and hexaploid wheat genomes, respectively [128]. When compared with the Illumina 9000 Infinium iSelect SNP assay [129], ddRADseq provided similar genome coverage and QTL detection at one-fourth of the cost [130]. Unsurprisingly, this cost-effective, high throughput approach is becoming widely adopted for genetic mapping of other large, polyploid genomes such as octoploid strawberry [131] and autotetraploid alfalfa [132].

RADseq has been accompanied by a suite of software pipelines to align, SNP call and identify polymorphic loci imputable in secondary mapping or phylogenetic software [133]. Stacks [134] and UNEAK [121] have emerged as the predominantly used pipelines for non-model species lacking a reference genome. Efforts such as the Sliding Window Extraction of Explicit Polymorphisms (SWEEP) pipeline have been made to identify

within sub-genome, homologous SNPs in peanut, reducing the rate of false-positive, homeologous SNP identification [135]. Many computational challenges remain in processing NGS data from complex non-model species [133] and as sequencing costs becomes economically inconsequential the bottleneck in genotyping such species is likely to occur at this stage [136].

Plant genotyping methods continue to become increasingly sophisticated, however, the utility of these approaches is only realized when *R* is significantly increased as a consequence [137]. Decades of breeding and genetics in staple crops has facilitated tremendous progress in connecting phenotype with genotype. Arguably the best example of this is in rice where marker assisted selection (MAS) facilitated submergent-tolerant marker *Sub1* introgression into widely cultivated variety 'Swarna' increasing yields 2-fold or higher throughout South and Southeast Asia [138]. Multiple marker assisted backcross (MABC) programs have also been employed for improved disease resistance including the pyramiding of *Pi-b* and *Pi-kh* rice blast resistance genes in the background of elite cultivar MR219 [139]. While successful in rice, QTL pyramiding for durable disease resistance has proven less tractable in other cases such as in downy mildew resistance in lettuce where epistasis was shown to complicate additive accumulation of resistance by stacking QTL in backcross introgression lines [110]. MAS has clearly been valuable for traits controlled by major genes or QTL as a function of greater *R* per introgressed genomic segment [137]. However, substantial limitations remain in more complex traits controlled by many genes represented by less reproducible, smaller or minor effect QTL [102].

The entire modern genetics and early genomic eras provided little insight into the genetics of non-model species due largely to prohibitive costs for reagents and sequencing [140]. NGS has lifted this burden, allowing for tremendous amounts of sequence data to be generated with modest funding. Current research of non-model species has the potential to be accelerated relative to their well-studied counterparts by avoiding mistaken paths and exploiting the many tools generated from the study of model species [136]. The magnitude of genotyping and bioinformatics bottlenecks initially experienced by model species is exponentially receding [123,141]. Modified NGS approaches such as RADseq have revolutionized marker development in non-model species by removing the requirement of *a priori* genomic data and facilitating affordable population level mapping and phylogenetic studies [120,126]. It is apparent from existing studies in major crop species that initial efforts towards identification of markers associated with traits of interest should be focused on those under major gene control following the model set for rice. The *Sub1* locus was converted to a selectable marker from a single major QTL with a LOD score of 36.0 and PVE of 69% and successfully implemented in MAS. This QTL is now found in rice varieties grown on >15 million hectares [138]. Given a saturated genetic map coupled with precise phenotyping, replication of the major gene *Sub1* MAS model is now a feasible goal in less well-characterized plant genomes.

Success stories with practical application are emerging in 'orphan crops', which are grown primarily on marginal land in developing nations and have been neglected from genetic study and breeding efforts [140]. In chickpea, a RADseq approach was used to identify a 14 cM genomic region responsible for 58% of variance in drought response

[142]. Using MABC, the drought tolerance locus was introgressed into major variety JG 11 and other breeding lines grown in environments prone to high precipitation [143]. Similar gains have been made in the development of disease resistant breeding lines including Fusarium wilt resistant chickpea and rust resistant groundnut [143].

The traction gained from QTL discovery of major-gene-controlled traits is unlikely to translate to more complex traits controlled by minor effect loci in non-model plant species. Documented failures of QTL mapping to inform effective MAS for complex traits in cereal crops [144] made obvious the need for alternative approaches. Among the two most popular alternative genomic assisted breeding approaches to emerge are GWAS (or association mapping) and genomic selection (GS). GWAS examines the genomes of a set of individuals with unrelated pedigrees to exploit a comparatively high number of recombination events to identify co-segregation of alleles with a trait of interest [145]. This linkage disequilibrium (LD)-based approach yields greater (Kbp) resolution as compared to the Mbp resolution of QTL mapping providing greater proximity of a genome sequence to a target trait [145]. A major issue with MAS strategies such as gene pyramiding or MABC that rely on QTL derived from a single bi-parental population is the decrease or loss of QTL effect when introgressed in genome backgrounds of unrelated elite lines [110]. GWAS can be powerful when combined with QTL mapping when used with multi-parental families such as NAM, three-/four-way or diallel crosses [145]. This approach was used in six Arabidopsis RIL populations derived from crosses between 12 diverse parents to resolve map six candidate genes resistant to five strains of *H. arabidopsis* [87].

Although first proposed in 2001 by Meuweissen et al. [146], GS has only recently been employed in a meaningful way to plant systems [147]. This approach develops a prediction model for estimating the value of non-phenotyped individuals and their putative progeny from an initial training population of similar genetic composition [102]. GS has the advantage of requiring less individuals to be phenotyped and flexibility in experimental design, but is highly dependent upon abundance of genomic data [148]. The potential of GS is likely to be realized when WGS inevitably reaches a miniscule cost that facilitates re-sequencing of individual genomes in populations to generate a genomic profiles [148]. In the interim RADseq methods have been successfully employed in GS strategies even in complex genomes such as wheat [102] and *Miscanthus sinensis* [149]. Given that highly quantitative, complex traits such as yield are likely to be influenced in some way by loci across the entirety of the genome, GS is an appropriate strategy for movement of many favorable alleles into a single individual's genome. Still current prediction accuracies of GS typically do not exceed 60% and will require improved modeling based on training populations [147,148]. As with initial MAS efforts, non-model species may be once again best positioned to refrain from investing resources in GS until the cost to benefit ratio associated with this approach is favorable [147].

Tremendous gains have been realized through the convergence of the genomic tools with a rich germplasm collected over the course of a millennium in some staple crops. Despite advances in DNA-based approaches plant breeding remains fundamentally unchanged as a means by which to congregate favorable alleles into a single genome to improve a trait or traits of interest. Diverse germplasm therefore remains foundational to effective breeding programs and a potential limiting factor for *R*. Effectiveness of

molecular breeding approaches such as QTL mapping, GWAS and GS strategies are highly dependent upon well-conceived populations from germplasm of known genetic relatedness [145,148,150]. Comprehensive collection and phylogenetic characterization is an important investment necessary to assess population structure extant in available germplasm [141]. Molecular classification of relationships among a germplasm panel is quite accessible even for nascent breeding programs of non-model species using a modest number (15-25) of informative SSR markers [151].

While breeding efforts in non-model species are fortunate to inherent vastly improved NGS-based genotyping, researchers should anticipate the phenotyping bottleneck being faced by those working in model plant systems [152]. Genomic data in MAS strategies are only as good as their predictive ability for a target trait in a given environment. This concept has become exceptionally important in recent years and even led to the coining of the term ‘phenomics’ [152]. While DNA sequence data is static, phenotype data is highly dynamic and subject to extensive variation as a function of interaction with the environment. Development of informative and scalable phenotype variation capture is technically challenging, laborious and costly [152]. The most substantial progress has been made in field crops where high-throughput canopy phenotyping using aerial sensor systems correlates leaf absorption at specific wavelengths in the electromagnetic spectrum to traits of interest [153]. Such non-invasive measurements are highly advantageous in that they allow for data point collection at multiple life cycle stages. This technology has shown tremendous progress but is cost-prohibitive to most plant breeding programs. Phenomics research is therefore in its infancy and sophisticated field and greenhouse sensor-based phenotyping is likely to

remain only economically feasible for major crops. Furthermore, in species such as those belonging to ‘orphan crops’ target traits such as nutrient and micronutrient content will present unique phenotyping challenges that may require custom systems or protocols. Nevertheless as genomic resources become increasingly available and ultimately plateau, increased precision and accuracy of phenotyping will fast become the limiting factor in non-model species as well. Early investment in optimization of phenotyping systems is likely to be a worthy use of resources to a non-model plant breeding program to mediate the inevitable phenomics bottleneck.

In conclusion, the successes and failures experience in breeding cereal and other major crop species provide an advantage to non-model species. Furthermore, it is fair to hypothesize that specialty crops are now poised to enter the genomic era given the accessibility of NGS and genotyping methodologies tailored to poorly characterized genomes. Sweet basil is a poorly characterized tetraploid species with no reference genome and therefore an excellent candidate to test this hypothesis. Successful execution of modern plant breeding techniques in this system will better equip this plant species to respond to future stresses and demonstrate the feasibility of these approached for other non-model species.

Thesis Overview

In the dissertation presented herein results from a series of studies that collectively outline a strategy for identification, characterization and introgression of resistance to a devastating disease downy mildew (*Peronospora belbahrii*) in a non-model genetic system, sweet basil (*Ocimum basilicum*). The major objectives of this research were to provide a framework for the 1) fusion of applied and basic studies to

inform realized genetic improvement with respect to downy mildew resistance and 2) development of a platform to effectively breed a poorly characterized plant species (*O. basilicum*). In chapter 1, a core set of cross-species transferable expressed sequence tag simple sequence repeat (EST-SSR) markers were developed and used to facilitate the most extensive assessment of population structure and genetic diversity of *Ocimum* spp. to date. Evaluation of all accessions for response to downy mildew provided insight into the distribution of resistant genotypes among heterotic *Ocimum* spp. populations. Results of this study informed the selection of parents MRI and SB22 for the development of a full-sibling family subsequently evaluated for response to downy mildew over two years and two locations to determine mode of inheritance for resistance conferred by genotype MRI. Finally, the MRI x SB22 F_2 mapping population was used to construct the first genetic map of sweet basil using double digestion restriction site associated DNA sequencing (RADseq). QTL analysis was performed and single nucleotide polymorphism (SNP) markers with significant association to MRI-conferred downy mildew resistance were identified. It is expected that response to future biotic and biotic stresses of sweet basil will be more quickly addressed using the information and strategies provided in this dissertation. Furthermore, results of this research demonstrate that breeding and genetics in a non-model plant system can prove fruitful given adherence to classical genetic theory in combination with modern genomic technologies.

References

1. Simon, J. E.; Quinn, J.; Murray JG. Basil: A Source of Essential Oils. In: Janick, J.; Simon JE, editors. Advances in New Crops. Portland, OR: Timber Press; 1990. pp. 484–489.
2. Wyenandt CA, Simon JE, Pyne RM, Homa K, McGrath MT, Zhang S, et al. Basil Downy Mildew (*Peronospora belbahrii*): Discoveries and Challenges Relative to Its Control. Phytopathology. 2015;105: 885–94. doi:10.1094/PHYTO-02-15-0032-FI
3. Putievsky E, Galambosi B. Production systems of sweet basil. In: Hiltunen R, Holm Y, editors. Basil: The Genus *Ocimum*. Amsterdam; 1999. pp. 39–61.
4. USDA. 2012. National Agricultural Statistics Service (NASS) Census of Agriculture. https://www.nass.usda.gov. (Accessed December13, 2016).
5. Gilardi G, Demarchi S, Garibaldi A, Gullino ML. Management of downy mildew of sweet basil (*Ocimum basilicum*) caused by *Peronospora belbahrii* by means of resistance inducers, fungicides, biocontrol agents and natural products. Phytoparasitica. 2013;41: 59–72. doi:10.1007/s12600-012-0264-y
6. Elad Y, Omer C, Nisan Z, Harari D, Goren H, Adler U, et al. Passive heat treatment of sweet basil crops suppresses *Peronospora belbahrii* downy mildew. Ann Appl Biol. 2016;168: 373–389. doi:10.1111/aab.12269
7. Koroch AR, Juliani HR, Sims C, Simon JE. Antioxidant activity, total phenolics, and rosmarinic acid content in different basils (*Ocimum* spp.). Isr J Plant Sci. 2010; 58: 191–195. doi:10.1560/IJPS.58.3-4.191
8. Jayasinghe C, Gotoh N, Aoki T, Wada S. Phenolics composition and antioxidant activity of sweet basil (*Ocimum basilicum* L.). J Agric Food Chem. 2003;51: 4442–4449. doi:10.1021/jf034269o
9. Swarup V, Ghosh J, Ghosh S, Saxena A, Basu A. Antiviral and anti-inflammatory effects of rosmarinic acid in an experimental murine model of Japanese encephalitis. Antimicrob Agents Chemother. 2007;51: 3367–3370. doi:10.1128/AAC.00041-07
10. Pichersky E, Gershenzon J. The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. Curr Opin Plant Biol. 2002;5: 237–243. doi:10.1016/S1369-5266(02)00251-0
11. Chiang LC, Ng LT, Cheng PW, Chiang W, Lin CC. Antiviral activities of extracts and selected pure constituents of *Ocimum basilicum*. Clin Exp Pharmacol Physiol. 2005;32: 811–816. doi:10.1111/j.1440-1681.2005.04270.x
12. Suppakul P, Miltz J, Sonneveld K, Bigger SW. Antimicrobial properties of basil and its possible application in food packaging. J Agric Food Chem. 2003;51: 3197–3207. doi:10.1021/jf021038t
13. Edris AE, Farrag ES. Antifungal activity of peppermint and sweet basil essential oils and their major aroma constituents on some plant pathogenic fungi from the vapor phase. Food / Nahrung. WILEY-VCH Verlag GmbH; 2003;47: 117–121. doi:10.1002/food.200390021
14. Grayer RJ, Kite GC, Goldstone FJ, Bryan SE, Paton A, Putievsky E. Infraspecific taxonomy and essential oil chemotypes in sweet basil, *Ocimum basilicum*.

Phytochemistry. 1996;43: 1033–1039. doi:10.1016/S0031-9422(96)00429-3
15. Carović-Stanko K, Liber Z, Politeo O, Strikić F, Kolak I, Milos M, et al. Molecular and chemical characterization of the most widespread *Ocimum* species. Plant Syst Evol. 2011;294: 253–262. doi:10.1007/s00606-011-0471-x
16. Vieira RF, Grayer RJ, Paton A, Simon JE. Genetic diversity of *Ocimum gratissimum* L. based on volatile oil constituents, flavonoids and RAPD markers. Biochem Syst Ecol. 2001;29: 287–304. doi:10.1016/S0305-1978(00)00062-4
17. Marotti M, Piccaglia R, Giovanelli E. Differences in Essential Oil Composition of Basil (*Ocimum basilicum* L.) Italian Cultivars Related to Corphological characteristics. J Agric Food Chem. 1996;44: 3926–3929. doi:10.1021/jf9601067
18. Nation RG, Janick J, Simon JE. Estimation of outcrossing in basil. Hortic Sci. 1992; 27: 1221–1222.
19. Simon JE, Quinn J, Murray J. Basil: A Source of Essential Oils. In: Janick J, Simon JE, editors. Advances in New Crops. Portland, OR: Timber Press; 1990. pp. 484–489.
20. Paton A, Harley MR, Harley MM. *Ocimum*: An overview of classification and relationships. Basil: The Genus *Ocimu*m. Amsterdam: Harwood Academic Publishers; 1999. pp. 1–38.
21. Vieira RF, Simon JE. Chemical characterization of basil (*Ocimum* spp.) based on volatile oils. Flavour Fragr J. 2006;21: 214–221. doi:10.1002/ffj.1513
22. Paton AJ, Springate D, Suddee S, Otieno D, Grayer RJ, Harley MM, et al. Phylogeny and evolution of basils and allies (*Ocimeae*, *Labiatae*) based on three plastid DNA regions. Mol Phylogenet Evol. 2004;31: 277–299. doi:10.1016/j.ympev.2003.08.002
23. Vieira RF, Goldsbrough P, Simon JE. Genetic diversity of basil (*Ocimum* spp.) based on RAPD markers. J Am Soc Hort Sci. 2003;128: 94–99.
24. Labra M, Miele M, Ledda B, Grassi F, Mazzei M, Sala F. Morphological characterization, essential oil composition and DNA genotyping of *Ocimum basilicum* L. cultivars. Plant Sci. 2004; 167:725–731. doi:10.1016/j.plantsci.2004.04.026
25. Khosla MK, Tawi- J, Group S. Karyomorphological Studies in Genus *Ocimum*. Cytologia. 1985;50: 253–263.
26. Paton AJ, Putievesky E. Taxonomic problems and cytotaxonomic relationships between and within varieties of *Ocimum basilicum* and related species (*Labiateae*). Kew Bull. 1996;51: 509–524.
27. Pushpangandan P, Sobti SN. Cytogenetical studies in the genus *Ocimum*. I. Origin of *O. americanum*, cytotaxonomical and experimental proof. Cytologia. 1982;47: 575–583.
28. Carović-Stanko K, Liber Z, Besendorfer V, Javornik B, Bohanec B, Kolak I, et al. Genetic relations among basil taxa (*Ocimum* L.) based on molecular markers, nuclear DNA content, and chromosome number. Plant Syst Evol. 2010;285: 13–22. doi:10.1007/s00606-009-0251-z
29. Koroch AR, Wang W, Michael TP, Dudai N, Simon JE, Belanger FC. Estimation of nuclear DNA content of cultivated *Ocimum* species by using flow cytometry. Isr J Plant Sci. 2010; 58: 183–189. doi:10.1560/IJPS.59.3-4.183
30. Putievesky E, Paton A, Lewinsohn E, Ravid U, Haimovich D, Katzir I, et al.

Crossability and relationship between *Ocimum basilicum* L. J Herbs, Spices, Med Plants. 1999;6: 11–24.
31. P.T.N. Spencer-Phillips; U. Gisis; A. Lebeda, editor. Advances in Downy Mildew Research. Kluwer Academic Publishers; 2002.
32. Beakes GW, Glockling SL, Sekimoto S. The evolutionary phylogeny of the oomycete "fungi." Protoplasma. 2012;249: 3–19. doi:10.1007/s00709-011-0269-2
33. Agrios GN. Plant Pathology. 5th ed. Burlington, MA: Elsevier Academic Press; 2005. doi:10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C
34. Thines M, Choi Y. Evolution, Diversity, and Taxonomy of the Peronosporaceae, with Focus on the Genus Peronospora. Phytopathology. 2015;106: 1–13. doi:10.1094/PHYTO-05-15-0127-RVW
35. de Bary A. Reserches sur le developpement de quelques champignons parasites. Ann Sci Nat Bot Ser. 1863;20: 5–148.
36. Yerkes WD, Shaw CG. Taxonomy of the *Peronospora* species on Cruciferae and Chenopodiaceae. Phytopathology. 1959;49: 499–507.
37. Choi YJ, Hong SB, Shin HD. Re-consideration of *Peronospora farinosa* infecting *Spinacia oleracea* as distinct species, *Peronospora effusa*. Mycol Res. 2007; 111: 381–391. doi:10.1016/j.mycres.2007.02.003
38. Savory EA, Granke LL, Quesada-Ocampo LM, Varbanova M, Hausbeck MK, Day B. The cucurbit downy mildew pathogen *Pseudoperonospora cubensis*. Mol Plant Pathol. 2011; 12: 217–226. doi:10.1111/j.1364-3703.2010.00670.x
39. Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: Waltz with plants. Curr Opin Genet Dev. Elsevier Ltd; 2015; 35: 57–65. doi:10.1016/j.gde.2015.09.001
40. Voglmayr H. Phylogenetic relationships of Peronospora and related genera based on nuclear ribosomal ITS sequences. Mycol Res. 2003;107: 1132–1142. doi:10.1017/S0953756203008438
41. Belbahri L, Calmin G, Pawlowski J, Lefort F. Phylogenetic analysis and real time PCR detection of a presumably undescribed *Peronospora* species on sweet basil and sage. Mycol Res. 2005; 109:1276–1287. doi:10.1017/S0953756205003928
42. Thines M, Telle S, Ploch S, Runge F. Identity of the downy mildew pathogens of basil, coleus, and sage with implications for quarantine measures. Mycol Res. 2009;113: 532–540. doi:10.1016/j.mycres.2008.12.005
43. Hansford CG. Annual report of the mycologist. Rev Appl Mycol. 1933;12: 421–422.
44. Roberts PD, Raid RN, Harmon PF, Jordan SA, Palmateer AJ. First Report of Downy Mildew Caused by a *Peronospora* sp. on Basil in Florida and the United States. Plant Dis. 2009;93: 199. doi:10.1094/PDIS-93-2-0199B
45. McLeod A, Coertze S, Mostert L. First Report of a *Peronospora* Species on Sweet Basil in South Africa. Plant Dis. 2006;90: 1115. doi:10.1094/PD-90-1115A
46. Blomquist CL, Rooney-Latham S, Nolan PA. First Report of Downy Mildew on Field-Grown Sweet Basil Caused by a *Peronospora* sp. in San Diego County, California. Plant Dis. 2009;93: 968. doi:10.1094/PDIS-93-9-0968A
47. Kong XY, Wang S, Wan SL, Xiao CL, Luo F, Liu Y. First Report of Downy Mildew on Basil (*Ocimum basilicum*) in China. Plant Dis. 2015;99: 1642. doi:10.1094/PDIS-01-15-0077-PDN

48. Nagy G, Horváth A. Occurrence of Downy Mildew Caused by *Peronospora belbahrii* on Sweet Basil in Hungary. Plant Dis. 2011;95: 1034. doi:10.1094/PDIS-04-11-0329
49. Garibaldi A, Minuto G, Bertetti D, Gullino ML. Seed transmission of *Peronospora* sp. of Basil. J Plant Dis Prot. 2004;111: 465–469.
50. Inaba T, Takahashi K, Morinaka T. Seed transmission of spinach downy mildew. Plant Dis. 1983. 67: 1139–1141.
51. Cohen Y, Rubin AE, Galperin M, Ploch S, Runge F, Thines M. Seed transmission of *Pseudoperonospora cubensis*. PLoS One. 2014;9: 1-12. doi:10.1371/journal.pone.0109766
52. Djalali Farahani-Kofoet R, Römer P, Grosch R. Systemic spread of downy mildew in basil plants and detection of the pathogen in seed and plant samples. Mycol Prog. 2012;11: 961–966. doi:10.1007/s11557-012-0816-z
53. Cohen Y, Ben-Naim Y. Nocturnal fanning suppresses downy mildew epidemics in sweet basil. PLoS One. 2016;11: 1–15. doi:10.1371/journal.pone.0155330
54. Stassen JHM, Van den Ackerveken G. How do oomycete effectors interfere with plant life? Curr Opin Plant Biol. Elsevier Ltd; 2011;14: 407–414. doi:10.1016/j.pbi.2011.05.002
55. Ojiambo PS, Gent DH, Quesada-Ocampo LM, Hausbeck MK, Holmes GJ. Epidemiology and Population Biology of *Pseudoperonospora cubensis* : A Model System for Management of Downy Mildews. Annu Rev Phytopathol. 2015;53: 223–246. doi:10.1146/annurev-phyto-080614-120048
56. Neufeld KN, Ojiambo PS. Interactive effects of Temperature and leaf wetness duration on sporangia germination and infection of cucurbit hosts by *Pseudoperonospora cubensis*. Plant Dis. 2012; 96:345–353. doi:10.1094/PDIS-07-11-0560
57. Williams MG, Magarey PA, Sivasithamparam K. Effect of temperature and light intensity on early infection behaviour of a Western Australian isolate of *Plasmopara viticola*, the downy mildew pathogen of grapevine. Australas Plant Pathol. 2007;36: 325–331. doi:10.1071/AP07029
58. Djalali Farahani-Kofoet R, Romer P, Grosch R. Selecting basil genotypes with resistance against downy mildew. Sci Hortic (Amsterdam). Elsevier B.V.; 2014;179: 248–255. doi:10.1016/j.scienta.2014.09.036
59. Garibaldi A, Bertetti D, Gullino ML. Effect of leaf wetness duration and temperature on infection of downy mildew (*Peronospora* sp.) of basil. J Plant Dis Prot. 2007;114: 6–8.
60. Koroch AR, Villani TS, Pyne RM, Simon JE. Rapid Staining Method to Detect and Identify Downy Mildew (*Peronospora belbahrii*) in Basil. Appl Plant Sci. 2013;1: 1300032. doi:10.3732/apps.1300032
61. Cohen Y, Van den Langenberg KM, Wehner TC, Ojiambo PS, Hausbeck M, Quesada-Ocampo LM, et al. Resurgence of *Pseudoperonospora cubensis*: The Causal Agent of Cucurbit Downy Mildew. Phytopathology. 2015;105: 998–1012. doi:10.1093/pcp/pci095
62. Cohen Y, Vaknin M, Ben-Naim Y, Rubin AE, Galperin M, Silverman D, et al. First Report of the Occurrence and Resistance to Mefenoxam of *Peronospora belbahrii*, Causal Agent of Downy Mildew of Basil (*Ocimum basilicum*) in Israel.

Plant Dis. 2013;97: 692. doi:10.1094/PDIS-12-12-1126-PDN
63. Collina M, Merighi M, Turan C, Pirondi A, Minuto G, Brunelli A. First Report of Resistance of *Peronospora belbahrii*, Causal Agent of Downy Mildew of Basil, to Mefenoxam in Italy. Plant Dis. 2016;100: 1787. doi:10.1094/PDIS-02-16-0237-PDN
64. Pintore I, Gilardi G, Gullino ML, Garibaldi A. Detection of mefenoxam-resistant strains of *Peronospora belbahrii*, the causal agent of basil downy mildew, transmitted through infected seeds. Phytoparasitica. 2016; doi:10.1007/s12600-016-0538-x
65. Brown S, Koike ST, Ochoa OE, Laemmlen F, Michelmore RW. Insensitivity to the fungicide fosetyl-aluminum in California isolates of the lettuce downy mildew pathogen, *Bremia lactucae*. Plant Dis. 2004; 88: 502–508. doi:10.1094/PDIS.2004.88.5.502
66. Gill US, Lee S, Mysore KS. Host Versus Nonhost Resistance: Distinct Wars with Similar Arsenals. Phytopathology. 2015;105: 580–587. doi:10.1094/PHYTO-11-14-0298-RVW
67. Felix G, Duran JD, Volko S, Boller T. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. Plant J. 1999; 18: 265–276. doi:10.1046/j.1365-313X.1999.00265.x
68. Jones JDG, Dangl L. The plant immune system. Nature. 2006;444: 323–329. doi:10.1038/nature05286
69. Mysore KS, Ryu CM. Nonhost resistance: How much do we know? Trends Plant Sci. 2004;9: 97–104. doi:10.1016/j.tplants.2003.12.005
70. Flor HH. Inheritance of resistance to rust in flax. J Agric Res. 1947; 74: 241–262.
71. Flor HH. Inheritance of pathogenicity in *Melampsora lini*. Phytopathology. 1942; 32: 653–669.
72. Parker JE, Holub EB, Frost LN, Falk A, Gunn ND, Daniels MJ. Characterization of *eds1,* a mutation in Arabidopsis suppressing resistance to *Peronospora parasitica* specified by several different RPP genes. Plant Cell. 1996;8: 2033–46. doi:10.1105/tpc.8.11.2033
73. Feys BJ, Moisan LJ, Newman MA, Parker JE. Direct interaction between the *Arabidopsis* disease resistance signaling proteins, EDS1 and PAD4. EMBO J. 2001;20: 5400–5411. doi:10.1093/emboj/20.19.5400
74. Wiermer M, Feys BJ, Parker JE. Plant immunity: The EDS1 regulatory node. Curr Opin Plant Biol. 2005;8: 383–389. doi:10.1016/j.pbi.2005.05.010
75. Gassmann W, Hinsch ME, Staskawicz BJ. The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant J. 1999;20: 265–277. doi:10.1046/j.1365-313X.1999.00600.x
76. Bozkurt TO, Schornack S, Banfield MJ, Kamoun S. Oomycetes, effectors, and all that jazz. Curr Opin Plant Biol. Elsevier Ltd; 2012;15: 483–492. doi:10.1016/j.pbi.2012.03.008
77. Soanes DM, Talbot NJ. Moving targets: rapid evolution of oomycete effectors. Trends Microbiol. 2008;16: 507–510. doi:10.1016/j.tim.2008.08.002
78. McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. Genome Biol. 2006;7: 212. doi:10.1186/gb-2006-7-4-212
79. Lozano R, Hamblin MT, Prochnik S, Jannink J-L. Identification and distribution of

the NBS-LRR gene family in the Cassava genome. BMC Genomics. 2015;16: 360. doi:10.1186/s12864-015-1554-9

80. Arya P, Kumar G, Acharya V, Singh AK. Genome-wide identification and expression analysis of NBS-encoding genes in Malus x domestica and expansion of NBS genes family in Rosaceae. PLoS One. 2014;9: e107987. doi:10.1371/journal.pone.0107987
81. Parker JE, Coleman MJ, Szabò V, Frost LN, Schmidt R, van der Biezen EA, et al. The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the toll and interleukin-1 receptors with N and L6. Plant Cell. 1997;9: 879–894. doi:10.1105/tpc.9.6.879
82. McDowell JM, Dhandaydham M, Long TA, Aarts MG, Goff S, Holub EB, et al. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis*. Plant Cell. 1998;10: 1861–1874. doi:10.1105/tpc.10.11.1861
83. Botella MA, Parker JE, Frost LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, et al. Three genes of the *Arabidopsis* RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell. 1998;10: 1847–60.
84. McDowell JM, Cuzick A, Can C, Beynon J, Dangl JL, Holub EB. Downy mildew (*Peronospora parasitica*) resistance genes in Arabidopsis vary in functional requirements for *NDR1*, *EDS1*, *NPR1* and salicylic acid accumulation. Plant J. 2000;22: 523–529.
85. Bittner-Eddy, PD; Crute, IR; Holub, EB; Beynon, JL. *RPP13* is a simple locus in Arabidopsis thaliana for alleles that specify downy mildew resistance in *Peronospora parasitica*. Plant J. 2000;12: 177–188.
86. Sinapidou E, Williams K, Nott L, Bahkt S, Tör M, Crute I, et al. Two TIR:NB:LRR genes are required to specify resistance to *Peronospora parasitica* isolate Cala2 in *Arabidopsis*. Plant J. 2004;38: 898–909. doi:10.1111/j.1365-313X.2004.02099.x
87. Nemri A, Atwell S, Tarone a M, Huang YS, Zhao K, Studholme DJ, et al. Genome-wide survey of *Arabidopsis* natural variation in downy mildew resistance using combined association and linkage mapping. Proc Natl Acad Sci USA. 2010;107: 10302–10307. doi:10.1073/pnas.0913160107
88. Shen KA, Chin DB, Arroyo-Garcia R, Ochoa OE, Lavelle DO, Wroblewski T, et al. *Dm3* Is One Member of a Large Constitutively Expressed Family of Nucleotide Binding Site—Leucine-Rich Repeat Encoding Genes. Mol Plant-Microbe Interact. 2002;15: 251–261. doi:10.1094/MPMI.2002.15.3.251
89. Feechan A, Anderson C, Torregrosa L, Jermakow A, Mestre P, Wiedemann-Merdinoglu S, et al. Genetic dissection of a TIR-NB-LRR locus from the wild North American grapevine species Muscadinia rotundifolia identifies paralogous genes conferring resistance to major fungal and oomycete pathogens in cultivated grapevine. Plant J. 2013;76: 661–674. doi:10.1111/tpj.12327
90. Radwan O, Bouzidi MF, Mouzeyar S. Molecular characterization of two types of resistance in sunflower to Plasmopara halstedii, the causal agent of downy mildew. Phytopathology. 2011;101: 970–9. doi:10.1094/PHYTO-06-10-0163
91. Irish BM, Correll JC, Feng C, Bentley T, de Los Reyes BG. Characterization of a resistance locus (*Pfs-1*) to the spinach downy mildew pathogen (*Peronospora*

farinosa f. sp. *spinaciae*) and development of a molecular marker linked to *Pfs-1*. Phytopathology. 2008;98: 894–900. doi:10.1094/PHYTO-98-8-0894

92. Epinat C, Pitrat M. Inheritance of resistance of three lines of muskmelon (*Cucumis melo*) to downy mildew (*Pseudoperonospora cubensis*). Proc Cucurbitaceae: Evaluation and Enhancement of Cucurbit Germplasm. Charleston, SC; 1989. pp. 133–135.
93. Olczak-Woltman H, Marcinkowska J, Niemirowicz-Szczytt K. The genetic basis of resistance to downy mildew in *Cucumis* spp.-latest developments and prospects. J Appl Genet. 2011;52: 249–255. doi:10.1007/s13353-011-0030-8
94. Parra L, Maisonneuve B, Lebeda A, Schut J, Christopoulou M, Jeuken M, et al. Rationalization of genes for resistance to *Bremia lactucae* in lettuce. Euphytica. 2016;210: 309–326. doi:10.1007/s10681-016-1687-1
95. Van Damme M, Huibers RP, Elberse J, Van Den Ackerveken G. *Arabidopsis DMR6* encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. Plant J. 2008;54: 785–793. doi:10.1111/j.1365-313X.2008.03427.x
96. Barnes WC, Epps WM. An unreported type of resistance to cucumber downy mildew. Plant Dis Rptr. 1954;34: 620.
97. Cabral A, Stassen JHM, Seidl MF, Bautor J, Parker JE, van den Ackerveken G. Identification of *Hyaloperonospora arabidopsis* transcript sequences expressed during infection reveals isolate-specific effectors. PLoS One. 2011;6: 1–13. doi:10.1371/journal.pone.0019328
98. Holmes GJ, Ojambo PS, Hausbeck MK, Quesada-Ocampo L, P. AK. Resurgence of Cucurbit Downy Mildew in the United States: A Watershed Event for Research and Extension. 2014;99: 428-441.
99. Fukuoka S, Saka N, Mizukami Y, Koga H, Yamanouchi U, Yoshioka Y, et al. Gene pyramiding enhances durable blast disease resistance in rice. Sci Rep. 2015;5: 7773. doi:10.1038/srep07773
100. Vicente JG, Gunn ND, Bailey L, Pink DAC, Holub EB. Genetics of resistance to downy mildew in *Brassica oleracea* and breeding towards durable disease control for UK vegetable production. Plant Pathol. 2012;61: 600–609. doi:10.1111/j.1365-3059.2011.02539.x
101. Falconer DS, Mackay, Trudy FC. Introduction to Quantitative Genetics. 4th ed. NJ: Pearson; 1996.
102. Poland J, Rutkoski J. Advances and Challenges in Genomic Selection for Disease Resistance. Annu Rev Phytopathol. 2016;54: 79–98. doi:10.1146/annurev-phyto-080615-100056
103. Fehr WR. Principles of Cultivar Development. New York, NY: MacMillan Publishing Company; 1987.
104. Henning JA, Gent DH, Twomey MC, Townsend MS, Pitra NJ, Matthews PD. Precision QTL mapping of downy mildew resistance in hop (*Humulus lupulus* L.). Euphytica. 2015;202: 487–498. doi:10.1007/s10681-015-1356-9
105. Kozik EU, Klosinska U, Call AD, Wehner TC. Heritability and genetic variance estimates for resistance to downy mildew in cucumber accession ames 2354. Crop Sci. 2013;53: 177–182. doi:10.2135/cropsci2012.05.0297
106. Wang Y, VandenLangenberg K, Wehner TC, Kraan PAG, Suelmann J, Zheng X,

et al. QTL mapping for downy mildew resistance in cucumber inbred line WI7120 (PI 330628). Theor Appl Genet; 2016;7120. doi:10.1007/s00122-016-2719-x

107. Grube RC, Ochoa OE. Comparative genetic analysis of field resistance to downy mildew in the lettuce cultivars “Grand Rapids” and “Iceberg.” Euphytica. 2005;142: 205–215. doi:10.1007/s10681-005-1683-3
108. Simko I, Atallah AJ, Ochoa OE, Antonise R, Galeano CH, Truco MJ, et al. Identification of QTLs conferring resistance to downy mildew in legacy cultivars of lettuce. Sci Rep. 2013;3: 2875. doi:10.1038/srep02875
109. Jeuken M, Lindhout P. Lactuca saligna, a non-host for lettuce downy mildew (*Bremia lactucae*), harbors a new race-specific *Dm* gene and three QTLs for resistance. Theor Appl Genet. 2002;105: 384–391. doi:10.1007/s00122-002-0943-z
110. den Boer E, Pelgrom KTB, Zhang NW, Visser RGF, Niks RE, Jeuken MJW. Effects of stacked quantitative resistances to downy mildew in lettuce do not simply add up. Theor Appl Genet. 2014;127: 1805–1816. doi:10.1007/s00122-014-2342-7
111. Kang YJ, Lee T, Lee J, Shim S, Jeong H, Satyawan D, et al. Translational genomics for plant breeding with the genome sequence explosion. Plant Biotechnol J. 2016;14: 1057–1069. doi:10.1111/pbi.12449
112. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature; 2000;408: 796–815. doi:10.1038/35048692
113. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, et al. The genome of woodland strawberry (*Fragaria vesca*). Nat Genet. 2011;43: 109–116. doi:10.1038/ng.740
114. Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KMM, et al. 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. Cell. 2016;166: 481–491. doi:10.1016/j.cell.2016.05.063
115. Cronn R, Knaus BJ, Liston A, Maughan PJ, Parks M, Syring J V., et al. Targeted enrichment strategies for next-generation plant biology. Am J Bot. 2012;99: 291–311. doi:10.3732/ajb.1100356
116. Pérez-de-Castro AM, Vilanova S, Cañizares J, Pascual L, Blanca JM, Díez MJ, et al. Application of genomic tools in plant breeding. Curr Genomics. 2012;13: 179–95. doi:10.2174/138920212800543084
117. Bassil N V, Davis TM, Zhang H, Ficklin S, Mittmann M, Webster T, et al. Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry *Fragaria* × *ananassa*. BMC Genomics. 2015;16: 155. doi:10.1186/s12864-015-1310-1
118. Kale SM, Pardeshi VC, Kadoo NY, Ghorpade PB, Jana MM, Gupta VS. Development of genomic simple sequence repeat markers for linseed using next-generation sequencing technology. Mol Breed. 2012;30: 597–606. doi:10.1007/s11032-011-9648-9
119. Straub SCK, Fishbein M, Livshultz T, Foster Z, Parks M, Weitemier K, et al. Building a model: developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. BMC Genomics. 2011;12: 211. doi:10.1186/1471-2164-12-211

120. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17: 81–92. doi:10.1038/nrg.2015.28
121. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011;6: 1–10. doi:10.1371/journal.pone.0019379
122. Biesecker LG, Shianna K V, Mullikin JC. Exome sequencing: the expert view. Genome Biol. 2011;12: 128. doi:10.1186/gb-2011-12-9-128
123. Davey JW., Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12: 499–510. doi:10.1038/nrg3012
124. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3: e3376. doi:10.1371/journal.pone.0003376
125. Poland J, Rife T. Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome. 2012;5: 92–102. doi:10.3835/plantgenome2012.05.0005
126. He J, Zhao X, Laroche A, Lu Z-X, Liu H, Li Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci. 2014;5: 484. doi:10.3389/fpls.2014.00484
127. Kim C, Guo H, Kong W, Chandnani R, Shuang LS, Paterson AH. Application of genotyping by sequencing technology to a variety of crop breeding programs. Plant Sci. 2016;242: 14–22. doi:10.1016/j.plantsci.2015.04.016
128. Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One. 2012;7. doi:10.1371/journal.pone.0032253
129. Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Natl Acad Sci USA. 2013;110: 8057–8062. doi:10.1073/pnas.1217133110
130. Bajgain P, Rouse MN, Anderson JA. Comparing genotyping-by-sequencing and single nucleotide polymorphism chip genotyping for quantitative trait loci mapping in wheat. Crop Sci. 2016;56: 232–248. doi:10.2135/cropsci2015.06.0389
131. Davik J, Sargent DJ, Brurberg MB, Lien S, Kent M, Alsheikh M. A ddRAD based linkage map of the cultivated Strawberry, *Fragaria x ananassa*. PLoS One. 2015;10: 1–10. doi:10.1371/journal.pone.0137746
132. Li X, Wei Y, Acharya A, Jiang Q, Kang J, Brummer EC. A saturated genetic linkage map of autotetraploid alfalfa (*Medicago sativa* L.) developed using genotyping-by-sequencing is highly syntenous with the *Medicago truncatula* genome. G3 (Bethesda). 2014;4: 1971–9. doi:10.1534/g3.114.012245
133. Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA. Single nucleotide polymorphism identification in polyploids: A review, example, and recommendations. Mol Plant. 2015;8: 831–846. doi:10.1016/j.molp.2015.02.002
134. Catchen JM, Amores A., Hohenlohe P, Cresko W, Postlethwait JH, De Koning D-J. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. Genes|Genomes|Genetics. 2011;1: 171–182. doi:10.1534/g3.111.000240
135. Clevenger JP, Ozias-Akins P. SWEEP: A Tool for Filtering High-Quality SNPs in

Polyploid Crops. G3 Genes|Genomes|Genetics. 2015;5: 1797–1803. doi:10.1534/g3.115.019703

136. Armstead I, Huang L, Ravagnani A, Robson P, Ougham H. Bioinformatics in the orphan crops. Brief Bioinform. 2009;10: 645–653. doi:10.1093/bib/bbp036
137. De La Fuente GN, Frei UK, Lübberstedt T. Accelerating plant breeding. Trends Plant Sci. 2013;18: 667–672. doi:10.1016/j.tplants.2013.09.001
138. Septiningsih EM, Pamplona AM, Sanchez DL, Neeraja CN, Vergara G V., Heuer S, et al. Development of submergence-tolerant rice cultivars: The *Sub1* locus and beyond. Ann Bot. 2009;103: 151–160. doi:10.1093/aob/mcn206
139. Tanweer FA, Rafii MY, Sijam K, Rahim HA, Ahmed F, Ashkani S, et al. Introgression of Blast Resistance Genes (Putative *Pi-b* and *Pi-kh*) into Elite Rice Cultivar MR219 through Marker-Assisted Selection. Front Plant Sci. 2015;6: 1–11. doi:10.3389/fpls.2015.01002
140. Varshney RK, Ribaut JM, Buckler ES, Tuberosa R, Rafalski AJ, Langridge P. Can genomics boost productivity of orphan crops? Nat Biotechnol. 2012;30: 1172–1176. doi:10.1038/nbt.2440
141. Varshney RK, Singh VK, Hickey JM, Xun X, Marshall DF, Wang J, et al. Analytical and Decision Support Tools for Genomics-Assisted Breeding. Trends Plant Sci. Elsevier Ltd; 2015;xx: 1–10. doi:10.1016/j.tplants.2015.10.018
142. Jaganathan D, Thudi M, Kale S, Azam S, Roorkiwal M, Gaur PM, et al. Genotyping-by-sequencing based intra-specific genetic map refines a "QTL-hotspot" region for drought tolerance in chickpea. Mol Genet Genomics. 2015;290: 559–571. doi:10.1007/s00438-014-0932-3
143. Varshney RK. Exciting journey of 10 years from genomes to fields and markets: Some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. Plant Sci. 2016;242: 98–107. doi:10.1016/j.plantsci.2015.09.009
144. Bernardo R. Molecular Markers and Selection for Complex Traits in Plants : Learning from the Last 20 Years. 2008; 1649–1664. doi:10.2135/cropsci2008.03.0131
145. Gupta PK, Kulwal PL, Jaiswal V. Association mapping in crop plants: Opportunities and challenges. In: Advances in Genetics. Amsterdam: Elsevier; 2014 pp. 109-147. doi:10.1016/B978-0-12-800271-1.00002-0
146. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. 2001;57: 1819-1829.
147. Heslot N, Jannink J-L, Sorrells ME. Perspectives for Genomic Selection Applications and Research in Plants. Crop Sci. 2015;55: 1–12. doi:10.2135/cropsci2014.03.0249
148. Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, et al. Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding. Front Genet. 2016;7: 1–11. doi:10.3389/fgene.2016.00221
149. Slavov GT, Nipper R, Robson P, Farrar K, Allison GG, Bosch M, et al. Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus sinensis*. New Phytol. 2014;201: 1227–1239. doi:10.1111/nph.12621
150. Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, et al. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time

and pathogen resistance genes. PLoS Genet. 2005;1. doi:10.1371/journal.pgen.0010060

151. Urrestarazu J, Denancé C, Ravon E, Guyader A, Guisnel R, Feugey L, et al. Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. BMC Plant Biol. 2016;16: 1–20. doi:10.1186/s12870-016-0818-0
152. Furbank RT, Tester M. Phenomics - technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 2011;16: 635–644. doi:10.1016/j.tplants.2011.09.005
153. Pauli D, Chapman SC, Bart R, Topp CN, Lawrence-Dill CJ, Poland J, et al. The quest for understanding phenotypic variation via integrated approaches in the field environment. Plant Physiol. 2016;4: 622-634. doi:10.1104/pp.16.00592

CHAPTER 2.

Population structure, genetic diversity and downy mildew resistance among *Ocimum* spp. germplasm

ABSTRACT

The basil (*Ocimum* spp.) genus maintains a rich diversity of phenotypes and aromatic volatiles through natural and artificial outcrossing. Yet the characterization of population structure and genetic diversity for this genus is severely lacking. Absence of this information has impeded the development of genetic resistance to diseases such as downy mildew (DM) caused by obligate oomycete *Peronospora belbahrii* and now a worldwide epidemic. In an effort to address this deficit, 20 EST-SSR markers with species-level transferability were developed and used to resolve relationships among a diverse panel of 180 *Ocimum* spp. accessions with varying response to downy mildew. The objective was to identify population delineations within *Ocimum* spp. and classify relations among individuals using high-throughput, reproducible DNA markers (EST-SSRs) to inform breeding for the current *Peronospora belbahrii* outbreak and better mediate future biotic stresses.

The results obtained from multiple analyses including nested Bayesian model-based clustering, analysis of molecular variance (AMOVA), and unweighted pair group method using arithmetic average (UPGMA) analysis were synergized to provide an updated phylogeny of the *Ocimum* genus. Three (major) and seven (sub) population (cluster) models were identified and well supported (P<0.001) by PhiPT (Φ_{PT}) values of 0.433 and 0.344, respectively. EST-SSR markers provided allelic frequency data among clusters that support previously developed hypotheses with regard to allopolyploid

genome structure. Evidence of cryptic population structure was demonstrated (ΔK = 2, 5, 7) for the k1 *O. basilicum* cluster supporting the expectation of substantial gene flow among these accessions. UPGMA analysis provided best resolution of the 36-accession, DM resistant k3 cluster with consistently strong bootstrap support. The 90-accession k1 cluster was found to be highly susceptible to downy mildew with 'MRI' representing the only accession available for DM resistance breeding. Although a rich source of DM resistance, introgression of k3 into the commercially important k1 accessions is impeded by reproductive barriers as demonstrated by multiple sterile F1 hybrids included in this study. The k2 cluster located between k1 and k3, represents a viable source of introgressed resistance as evidenced by fertile backcross progeny.

High levels of genetic and genomic diversity support the observed phenotypic variation among *Ocimum* spp. accessions. EST-SSRs provide the most robust evaluation of molecular diversity to date and can be used for additional studies to increase resolution of the genus. Elucidation of population structure and genetic relationships among *Ocimum* spp. germplasm provide the foundation for developing more effective DM resistance breeding strategies and more rapidly respond to future outbreaks likely to occur given the propensity of DM race evolution documented in other crops. More broadly, results provide a basis for selection strategies that can continue to build and maintain genetic diversity in sweet basil.

Introduction

The genus *Ocimum* is estimated to contain approximately 64 herbaceous annual and perennial plant species [1]. The primary center of diversity is in Tropical Africa [2], while a secondary center exists in Tropical Asia and a tertiary in the New World tropics [2,3]. A number of species are cultivated and processed throughout the world for application in medicinal, dietary supplement, essential oil, food flavoring and culinary industries [4]. Sweet basil, *Ocimum basilicum*, is among the most economically important culinary herbs in the United States, Europe and Israel [5]. Plant breeding efforts have largely targeted sweet basil for improved disease resistance [6], chilling tolerance [7] and novel volatile profiles (chemotypes) [8].

Ocimum spp. demonstrate relatively high rates of outcrossing in the presence of pollinators [9], while readily self-pollinating in their absence. Importantly, recurrent self-pollination provides no evidence of inbreeding depression (data not provided). The flexibility of this reproductive system has been exploited by natural and artificial selection, likely serving as a catalyst for morphological and volatile diversification of *Ocimum* spp. [4]. In addition, cytological investigation suggests this genus has also undergone extensive genome augmentation resulting in ploidy variation. While the x = 12 basic chromosome is considered stable in tetraploid (2n=4x=48) *O. basilicum*, different ploidies have been reported across *Ocimum* spp. [10]. Dual basic chromosome numbers of 8 and 12 have been proposed for two major ‘Basilicum’ and ‘Sanctum’ groups, respectively [11]. More recently, studies involving nuclear DNA content support the hypothesis of x = 12 as the genus-wide basic chromosome number with some cases of aneuploidy [12,13]. This explanation accounts for dysploidy, which has proven to be a

persistent phenomenon in plant evolutionary biology[14]. The 386 Mbp *O. tenuiflorum (*syn. *O. sanctum*) genome was reported to be a diploid (2n=2x=16) supporting the existence of an x = 8 basic chromosome number and suggesting *O. tenuiflorum* is a diploid species [15]. Thus, competing theories of major genome structure remain prevalent in the *Ocimum* genus.

Examination of chromosome pairing behavior suggested allopolyploidy likely explains polyploid genomic structure in basil [10,11]. For instance, karyological investigation of *O. basilicum* and *O. basilicum* x *O. americanum* F_1 hybrids found no evidence of tetravalent and little trivalent formation (4-8% of pollen mother cells) demonstrating high levels of preferential pairing within subgenomes [11]. Relatively high levels of inter-specific hybridization [9] suggest allopolyploid formation to be more likely than autopolyploidy [16]. This hypothesis is also supported by genotype data from simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers in the *O. basilicum* MRI x SB22 F_2 mapping family demonstrated disomic segregation providing molecular evidence for an allotetraploid *O. basilicum* [17].

In addition to variation in ploidy [10,18], a variety of factors including unknown geographic origin and phenotypic diversity complicate classification of relationships among *O. basilicum* and the greater *Ocimum* genus. Taxonomical incongruences are also widespread throughout the literature and among seed distributors. Furthermore, absence of standardized descriptors or voucher specimens for particular cultivars has been particularly problematic in classifying *O. basilicum* [1]. Species of economic salience such as *O. x citriodorum*, *O. americanum* and *O. basilicum* var. *citriodorum* are cited interchangeably throughout the literature making comparison across studies challenging

[19]. Despite attempts to standardize nomenclature [19,20] the species epithets continue to be used redundantly (USDA GRIN).

Morphological characters and volatile composition (chemotype) have been used to estimate genetic diversity in the *Ocimum* genus [21–24]. Classification based on phenotypic characters is problematic in that homologous similarities may be difficult or impossible to discriminate from traits resulting from convergent evolution [25]. Reliance upon analyses using these metrics may be particularly damaging in the context of plant breeding, which relies heavily upon genetic relationships among germplasm to make informed decisions with regard to selection and crossing decisions [26]. Nevertheless, the most comprehensive study of *Ocimum* to date remains that of Paton et al [1], which relied exclusively upon morphological characters [1]. Results of this study delineated three major divisions in the genus or subgenera corresponding to *Ocimum*, *Nautochilus* and *Gymnocium* [1]. Subgenera *Ocimum* is further divided into three sections including *Gratissima*, *Hiantia*, and *Ocimum*, which *O. basilicum* and *O. americanum* were placed [1].

Molecular phylogenetic approaches have employed RAPD and/or AFLP markers to determine relationships among nine [27], 12 [28], 22 [18], 28 [12] and 37 [2] *Ocimum* spp. accessions. DNA markers from three plastid regions were used by Paton et al. [3] to classify 12 *Ocimum* spp. accessions within the Lamiaceae family [3]. More recently, inter-simple sequence repeat (ISSR) markers were found to have higher polymorphism information content (PIC) and greater resolving power than chloroplast DNA (cpDNA) *psbA-trnH* markers following cluster analysis of 12 accessions representing 8 *Ocimum* spp.[29]. Despite being the preferred marker systems for phylogenetic analyses SSR

markers have not been used to classify relationships among *Ocimum* spp. accessions. A recent study validation of six EST-SSR markers across six accessions, however, phylogenetic characterization of the genus remains to be seen [30]. SSRs are considered advantageous because they are multi-allelic, polymorphic, reproducible, ubiquitous throughout the genome and amenable to high-throughput genotyping with multiplexed PCR [31]. Expressed sequence tag (EST) derived SSRs (EST-SSRs) are particularly desirable for genetic diversity and mapping studies due to their location in coding regions of the genome and transferability across related germplasm [32,33]. The National Center for Biotechnology Information (NCBI) *O. basilicum* EST library was recently used to develop and genotype the MRI x SB22 mapping population [17].

The deficit of available sweet basil genetic and genomic resources relative to other cultivated plant species has been highlighted in recent years by a worldwide downy mildew (DM) epidemic caused by the oomycete pathogen *Peronospora belbahrii* [34,35]. The deployment of genetic resistance is an essential management strategy for effective disease control as demonstrated by extensive breeding across plant species affected by DM [35–39]. Recent reports of Mefenoxam resistant *P. belbahrii* isolates in Israel [40] and Italy [41,42] exacerbate the need for resistant cultivars. However, development of resistant sweet basil varieties has been dramatically slowed by a severe lack of knowledge regarding genetic similarity between elite *O. basilicum* and exotic *Ocimum* spp. germplasm rich in sources of resistance [43–45]. This has rendered attempted introgression of resistant loci into sweet basil accessions subject to rudimentary trial and error by laborious cross-pollinations between elite and exotic breeding lines. Among 27

F_1 hybrids evaluated by Ben-Naim et al. [43], those considered 'highly resistant' were completely sterile [43].

Population structure is among the analyses most needed to better elucidate relations among *Ocimum* spp. Model-based clustering is a powerful analysis method frequently used to infer association of individuals to distinct populations from multilocus data sets [46]. Association of germplasm to distinct populations provides essential *a priori* information needed to avoid issues of sterility while monitoring potential narrowing of the genetic pool among breeding populations. Furthermore, detection of cryptic population structure is an important prerequisite to association mapping (AM) of alleles with traits of interest (i.e., disease resistance) by avoiding false positive associations [46,47]. Identification of population structure and estimation of genetic diversity among *Ocimum* spp. is needed to provide a more robust breeding response to future biotic or abiotic stresses.

In this study no *a priori* data or designations of species/relationships are considered to provide an unbiased update of basil phylogeny and first insights into population structure. The objectives of this study were to (i) develop a reliable set of EST-SSR markers capable of resolving genetic relationships among basil accessions, (ii) determine population structure and genetic diversity among a diverse panel of 180 *Ocimum* spp. accessions, and (iii) determine the distribution of downy mildew resistance among a representative *Ocimum* spp. germplasm panel.

Materials and Methods

Plant Material

A 180-accession panel of *Ocimum* spp. was selected to provide a genus-wide representation of genetic diversity, while targeting germplasm developed primarily for improved performance in important agronomic traits at Rutgers University. Eight chilling ('RUCB') and 7 *Fusarium oxysporum* f.sp. *basilici* (FOB) ('RUSB') tolerant *O. basilicum* inbred lines were selected. Twenty-nine breeding lines were included from two lineages ('RU4S' and 'RUMS') selected for DM resistance. Eighteen F_1 hybrids were developed from hybridization of 9 *Ocimum* spp. accessions of varying DM response with 4 DM-susceptible *O. basilicum* inbred lines. Seven F_2 selections, a single $F_{2:3}$ individual and four first-generation backcross lines selected for FOB resistance ('RU172S'). All cross-pollinations were performed as described by Pyne et al. [56] and self-pollination was performed by single plant isolation.

Remaining accessions were obtained from either commercial seed companies or the USDA-GRIN, representing 10 species according to these sources. Finally, two catnip (*Nepeta cataria*) cultivars CN3 and CR9 were included as outgroup species.

Disease Rating (Phenotypic Evaluation)

Seed sourced from commercial, USDA-GRIN and inbred accessions were planted in Fafard Growing Mix 2 (Sun Gro Horticulture, Agawam, MA) and germinated under intermittent misting. Hybrid and non-flowering accessions ('Pesto Perpetuo', 'GCB' and 'Pezou') were cloned by vegetative cutting from mother plants. A *P. belbahrii* isolate collected at the Rutgers Agricultural Research and Extension Center (RAREC) in 2013 and maintained on susceptible check variety 'DiGenova' was used in this study. Inoculum was prepared and applied to first true leaf sets according to Pyne et al. [39]. Eight seedlings (individuals) per each accession were inoculated and a disease ratings

were performed 10 days post inoculation (DPI) corresponding to the interval at which disease severity was greatest. Individual plants were assigned a disease severity (DS) score using an ordered categorical scale in which 0=no sporulation, 1=1-10%, 2=11-25%, 3=26-50%, 4=51-100% [56]. A disease severity score was assigned to each accession from the average of 8 individuals for two, repeated experiments.

EST-SSR Genotyping

Young leaves were harvested from axillary nodes of each accession in this study, frozen and ground. Genomic DNA (gDNA) was extracted from ~80 mg of ground leaf tissue for all accessions using the E.N.Z.A. SP Plant DNA Kit (Omega BioTek, Norcross, GA).

A set of 240 EST-SSR markers were previously developed and used to genotype the accessions MRI and SB22 [17]. Preliminary experiments identified 20 di-, tri- and tetranucleotide repeat markers demonstrating reliable, unambiguous amplification across replicated PCR evaluations in a diverse subset of *Ocimum* spp. accessions. Given the putative allopolyploid genome structure of these species, markers were considered to be amplified in shared sub-genomes and, thus, orthologous. Seven of the 20 selected markers were previously mapped to sweet basil linkage map developed from MRI x SB22 F_2 mapping population [17]. All primers were synthesized by Integrated DNA technologies (Coralville, IA). Multiplexed PCR was performed using conditions and fluorescent dyes (FAM, NED, PET and VIC) previously described [17]. Fragments generated from PCR were separated by capillary electrophoresis using an ABI 3500xL Genetic Analyzer (Life Technologies Corporation, Carlsbad, CA). Genemapper 4.1

(Applied Biosystems) software was used to determine fragment size for alleles across all accessions and exported for downstream analysis.

Polyploidy is a confounding factor in population genetic studies due to the complication of allelic configuration among genotypes. In allopolyploid accessions of a single ploidy amplification of loci in independent sub-genomes provides a platform for conversion to a diploid SSR genotyping system [50,63]. Accurate, multi-allelic (codominant) genotype assignment of complex germplasm panels with multiple ploidy levels in the absence of subgenome assignment exceeds current data analysis capabilities. Genotype data in this study was therefore scored as binary (presence = 1; absence/null = 0) as described by Honig et al. [64]. SSR markers absent of any alleles (primer site mutation) or having many (>6) alleles (non-specific primer binding) were considered null and coded '0'. Genotypes were coded '-1' in the case of missing data. Polymorphic information content (PIC) was calculated for individual alleles using the formula $2P_iQ_i$, where P_i is the frequency of presence and Q_i is the frequency of absence for a given allele [64].

Phenotype Data Analysis

Analysis of variance (ANOVA) was performed using R software to determine whether the effect of genotype was significant with respect to the response variable, DS. The 'agricolae' package in R was used to perform DS mean separation using Tukey's honest significant difference (HSD) test with $\alpha = 0.05$. All accessions were included in this analysis with the exception of PI 511865, which was found to segregate and, thus, could not be treated as a single genotype.

Population Structure Analysis

To determine population structure, a nested Bayesian model-based clustering approach described by Vähä et al. [65] was implemented with Structure ver 2.3.4 software [46] to infer assignment of individuals to *K* clusters and sub-clusters. This approach has been effective in identifying important major and cryptic population structure in various plant species including apple [66], cotton [50] and rice [67]. All simulations were performed using the 'admixture model' with 'correlated allele frequencies'. The algorithm was executed with parameters set to a burnin of 100,000 followed by 200,000 Markov Chain Monte Carlo (MCMC) repetitions with 10 replicated runs per *K*. Initial (primary) Structure analysis was performed for $K = 1 – 15$ using all 180 accessions to determine major delineations in overall population structure. The most parsimonious value of *K* was determined by Evanno's method [68] using Structure Harvester software v0.6.94 [69] to determine the natural logarithm probability Pr(X|K) and *ad hoc* ΔK statistic (mean(|L"(*K*)|) / sd(L(*K*)) from membership coefficient (*qI*) matrices. Following selection of the optimal *K* for all replicated runs, the FULLSEARCH algorithm of CLUMPP software [70] was used to identify the optimal permuted order of *qI* matrices with the greatest pairwise similarity. The average of this permutation provided the best *qI* matrix representation, which was visualized using DISTRUCT software [71].

Determinations of the *qI* threshold at which to infer cluster assignment vary by study, but range from $qI = 0.5$ [72] – 0.8 [66]. Inclusion of wide crosses (F_1, F_2 and BC_1) in this study provided a measure of high admixture and informed the selection of a 0.70 *qI* threshold. Thus, accessions were assigned to the cluster for which $qI > 0.70$, while those with a maximum *qI* less than 0.70 were considered admixed. In addition, hybrid

(F_1, F_2, $F_{2:3}$) and first backcross (BC_1) accessions with parentage assigned to two different clusters were considered admixed.

Nested (secondary) Structure analysis was performed separately for non-admixed accessions of each cluster ascertained from the primary analysis. Algorithm parameters were identical to that of the primary analysis but included K = 1-10. Hybrid and backcross accessions were then assigned to sub-clusters by replacing the primary *qI* with secondary *qI* determined for each parent. Accessions for which parentage was unknown were excluded from secondary Structure analysis. Assignment to sub-clusters was repeated using the same admixture criteria as in the primary analysis.

AMOVA

The integrity of clusters derived from population structure was investigated by performing an analysis of molecular variance (AMOVA) using GenAlEx 6.501 software [73]. A Nei [74] genetic distance matrix was first calculated using the binary genotype matrix previously described for the model-based clustering analyses. Two separate AMOVA were performed to partition within and among population estimated and molecular variance (σ^2) components. Statistical significance of pairwise genetic distance among (i) primary and (ii) secondary clustering iterations were used to calculate the pairwise population PhiPT (Φ_{PT}) test statistic, which provides a measure of interpopulation genetic diversity with intra-individual (heterozygosity) variation suppressed. The Φ_{PT} statistic was calculated for all pairwise cluster combinations, which were determined significantly different ($p<0.05$) by 999 random permutations of the data.

UPGMA

To investigate genetic relationships among accessions in the full panel, unweighted pair group method using arithmetic average (UPGMA) clustering was performed with Numerical Taxonomy System (NTSYSpc) ver 2.21q software (Exeter Software, Setauket, New York, USA) [48]. Genotype data from outgroup accessions CN3 and CR9 (*Nepata cataria*) were added to the same binary genotype matrix used for population structure and AMOVA analyses resulting in a total of 182 accessions. A genetic similarity matrix was generated using the Jaccard similarity coefficient method [75] in the NTSYSpc SIMQUAL module. Cluster analysis was then performed by UPGMA in the SAHN module and the output was visualized as a dendrogram with the TREE module. A Mantel test was performed with 999 test permutations using the MXCOMP module to determine goodness of fit between the genetic similarity matrix and the UPGMA dendrogram converted to cophenetic values with the COPH module. Finally, the original binary genotype matrix was resampled 1,000 times using the RESAMPLE module and the results were used as input for the CONSENS module to calculate bootstrap values using the majority rule method and a minimum support value of 0.500.

Results

Response to *Peronospora belbahrii*

Significant differences ($P<0.001$) were detected among 179 genotypes evaluated for response to DM (Table 1). Eighty-two accessions (46%) demonstrated a mean DS < 1.0 (considered resistant) of which 33 (18%) exhibited no sporulation (DS = 0). A large proportion (17%) of the accessions with mean DS < 1.0 included Rutgers breeding lines with MRI-conferred DM resistance (DM_{MRI}). DM susceptibility was widespread with 71

accessions (40%) having DS ≥ 3.0 (considered highly susceptible). The remaining 26 accessions (15%) were in an intermediate range between 1 (10%) and 3 (50%) representing levels of DM tolerance. A single accession (PI 511865) segregated for response to DM indicating heterozygosity among loci controlling resistance in this accession.

EST-SSR polymorphism

The 20 EST-SSR markers used in this study were pre-screened for unambiguous PCR amplification and demonstrated a low percentage of null genotypes. A minimum of 90% and an average of 96% of individuals were successfully genotyped across all EST-SSR markers. A total of 269 unique alleles were amplified across the 180 *Ocimum* spp. panel and two outgroup *Nepeta cateria* accessions (catnip cultivars CN3 and CR9). The average PIC across 20 EST-SSR markers was 0.14, ranging from 0.01 to 0.48 and 6 to 24 alleles per marker (locus) (Table 2). The average number of number of alleles/locus across the full panel was 1.86 (Table 3).

Population structure and allele frequency

Primary model-based clustering (population structure) analysis provided unambiguous evidence (Additional file 1. Figure S1) for three major clusters (K=3) k1 (n=90), k2 (n=16) and k3 (n=36) (Fig. 1A). The remaining 38 accessions were considered admixed due to qI<0.70 and/or genotypes of bi-parental origin from different clusters (Table 1). The average number of alleles per locus among clusters differed with values of 1.75, 2.23 and 1.34 corresponding to k1, k2 and k3, respectively (Table 3). Furthermore, clear differences were observed in the distribution of alleles per locus among clusters derived from primary population structure analysis (Fig. 2). The majority of loci in the k1

cluster had 2 alleles (57%), nearly twice the 33% observed for single-allele loci. A more even distribution was observed in the k2 cluster with 22%, 37% and 36% for one, two and three alleles per locus. This rate of tri-allelic loci is in stark contrast to the <1% in k1 and k3 clusters and suggests three homozygous, homeologous loci among a proportion of markers in k2 cluster. A majority of loci in k3 cluster were mono-allelic, twice that of the 30% bi-allelic loci. Finally, the number of alleles for loci of admixed accessions was relatively evenly distributed from 1 to 4 (Fig. 2) ranging from 18% to 32%. Five alleles were observed in an additional 4.3% of admixed loci. A total of 1812 alleles and an average of 2.58 alleles per locus (highest among all clusters) were detected for the 38 admixed accessions indicating heterozygosity.

Secondary (nested) population structure analysis provided evidence for seven clusters (K=7) among 142 non-admixed accessions (Fig. 2A). Two (K=2) and three (K=3) sub-clusters within k2 (k2.1 and k2.2) and k3 (k3.1, k3.2 and k3.3.), respectively, were strongly supported by the ΔK statistic (Additional file 1. Figure S1). Greatest support was provided for two (K=2) sub-clusters, k1.1 and k1.2, among the 90-accession k1 primary cluster containing commercial sweet basil accessions. However, evidence was also provided for five (K=5) and seven (K=7) sub-clusters (Fig. 1B), suggesting cryptic population structure extant within this most economically important grouping. Nested population structure could not be determined for nine accessions with admixed cluster membership from k1 and k3. Among these accessions, qI for either cluster (k1 or k3) failed to exceed 0.7 and could not be assigned due to unknown parentage (Fig. 2). An additional 12 accessions had qI<0.70 after being subjected to the nested population structure analysis and were considered admixed (Table 1).

The average (Table 3) and distribution (Fig. 2) of the alleles per locus for sub-clusters was similar to the major primary clusters of k1 and k2. The range of alleles per locus among k1, k2 and k3 sub-clusters were 1.74-1.76, 2.18-2.27 and 1.30-1.41, respectively, indicating a high level of consistency with regard to allelic diversity. Interestingly, a relatively high percentage (11.2%) of null alleles occurred among accessions of sub-cluster k3.2.

Molecular Variance

AMOVA was performed using 142 and 130 non-admixed accessions determined by primary and nested model-based cluster analysis, respectively. Inter-population differentiation was supported at the $P<0.001$ level for primary and nested population structure with Φ_{PT} values of 0.433 and 0.344, respectively (Table 4). The molecular variance due to differences within populations for nested clustering was 66% as compared to 57% from primary clustering. All pairwise Φ_{PT} values for (K=3) clusters were significantly different ($P<0.001$) (Table 5). The majority of pairwise Φ_{PT} values were highly significant ($P<0.001$) for the seven (K=7) clusters derived from nested analysis (Table 6). However, k3.1, k3.2 and k3.3 failed to demonstrate evidence for population differentiation based on AMOVA results (Table 6). These three sub-clusters contributed 51% of the total within population sum of squares (SS_{WP}) while representing only 5.4% of the accessions included in the AMOVA (data not shown). Thus, high within population genetic variation coupled with a low sample size may have lowered resolution for differentiation of these sub-clusters.

Characterization of genetic relationships

A cophenetic correlation value of r = 0.934 indicated strong goodness of fit [48] between the genetic similarity matrix used and the UPGMA dendrogram generated. Furthermore, the UPGMA dendrogram is in agreement with the placement of accessions for all major clusters (*K*=3) determined by population structure analysis (Fig. 3A-D). The k1 cluster demonstrated generally lower genetic diversity as evidenced by high Jaccard similarity coefficient values (Fig. 3A-B). Importantly, accessions within the k1 cluster largely grouped consistent with known pedigree information available for 61% of accessions. The k2 cluster divides, according to nested population structure, into two clades of comparable genetic diversity. Although delineation of k3 into three sub-clusters was not supported by AMOVA, high bootstrap support values provide evidence of distinct groupings with high statistical backing (Fig. 3C-D). UPGMA grouping among k1 accessions did not demonstrate the same level of bootstrap support. However, this was not unexpected considering the ambiguity of nested analysis results, which provided *adhoc* ΔK support for three population models (*K*=2, 5 and 7) (Fig. 1B) suggesting complex population structure.

The k1 cluster roots to a single major node (Fig. 3B) that can be further divided into three distinct groups (Fig 3A-B). This k1 group begins with (3) ('Caesar') and ends with 31 ('Lettuce Leaf Heirloom') (Fig 3A). Included are all accessions classified as *O. basilicum* such as the commercially important 19 ('Nufar') originally developed for fusarium wilt (FOB) resistance and widely cultivated worldwide. In addition, all chilling (RU_CB) and FOB (RU_SB) tolerant Rutgers breeding lines are included in this group (Fig. 3A). A second major k1 clade includes two lineages (RU_MS and RU_4S) (Fig. 3A-B) derived from breeding for DM_{MRI} resistance. The resistant genotype 47 (MRI) is

centrally located among accessions in these lineages. USDA-GRIN accession 60 (PI 197442) and all remaining k1 accessions (Fig. 3B) vary substantially in phenotype and aromatic volatile composition (data not published). Accessions in this k1 clade are primarily cultivated for specialty or ornamental markets with the exception of 69 ('Eleonora'), a variety marketed for the fresh market.

A k1 x k2 admixed, hybrid clade is supported by UPGMA analysis and can be observed at the top of Fig. 3C. This clade includes F_1, F_2 or $F_{2:3}$ progeny from a bi-parental cross of known origin and low or no fertility. Accessions 116 (PI 172996) and 115 ('Sweet Dani') were used as k2 parents of all progeny with the exception of 97 (RU_SB22 x LIME F_1) and 98 (PI_652060 x RU_SB17_F1).

The k2 cluster corresponds to a major clade, which is further divided into two groups (Fig. 3C). Nested population structure analysis (Fig. 1A) and bootstrap support values (0.504 and 0.744) (Fig 3C) are both in agreement with this sub-division. Accessions in this cluster are described by their sources as *O. basilicum*, *O. citriodorum* or *O. americanum* (Table 1). Three accessions cluster separately from k2 accessions and constitute a k2 x k3 hybrid grouping. Another k1 x k3 admixed clade is well-supported (0.938) and includes three F_1 hybrids derived from a cross between DM resistant k3 accession 139 ('Spice') and three different k1 accessions. Two additional USDA-GRIN accessions, 140 (PI 414201) and (141) PI 414203, are included in this clade and parentage is unknown.

Although the division of k3 into sub-clusters k3.1, k3.2 and k3.3 suggested by nested population structure analysis was not supported by the pairwise Φ_{PT} estimates (Table 6), UPGMA provides evidence of clear delineations within this major cluster (Fig.

3D). The k3.1 cluster includes 7 phenotypically indistinguishable accessions sourced from commercial seed companies and the USDA-GRIN. This cluster is highly supported (1.00) and evidently an autonomous population (Fig. 1; Fig. 3C). Another well-supported clade (0.997) is the k3.3 cluster (Fig 1.), which includes accessions 161 (PI 500945) – 171 (PI 500942). UPGMA analysis suggests this cluster represents a rather substantial juncture from the rest of the panel as it and subsequent clades are considered more basal than even the two *Nepeta cataria* outgroup accessions, catnip cultivars CN3 and CR9, according to UPGMA clustering. The k3.3 cluster is phenotypically homogenous and labeled *O. x africanum* according to the USDA-GRIN system. The k3.2 cluster includes basal most clades of the UPGMA dendrogram and the highest levels of genetic diversity based on Jaccard Similarity Coefficient genetic distance estimates shown as branch lengths in Fig. 3D. All accessions within this clade, supported by a 0.509 bootstrap value, are *O. gratissimum* according to the sources (Table 1). Within this clade bootstrap support values of 0.875, 0.977 and 0.899 support grouping of three more closely related pairs of accessions (Fig. 3D). Accession 182 (PI 511865) is labeled *O. selloi* groups (0.980) with the phenotypically similar 181 (‘Green Pepper Basil’) (Fig 3D). *O. gratissimum* and *O. tenuiflorum* accessions were used as outgroups in a previous AFLP-based assessment of basil phylogenetics [12], but did not include *O. selloi* found to be more basal in this study.

A number of accessions show membership (qI>0.70) to the k3.2 cluster yet are separated according to UPGMA analysis (Fig. 1A). This includes the phenotypically similar accessions 145 (PI 652052) and 149 (‘Camphor’), the former is classified *O. x africanum* by the USDA-GRIN, but more likely *O. kilimandsharicum* based on

morphology and as described by Ben-Naim et al. [43]. These accessions group within a clade supported by a bootstrap support value of 0.907 (Fig. 3D). A second, well-supported (0.867) clade included those with inferred ancestry to k3.2 (Fig. 1). These accessions are classified as *O. tenuiflorum* by commercial and USDA-GRIN sources (Table 1). Strong bootstrap support and phenotypic homogeneity among these accessions indicate UPGMA best resolves these k3 relationships.

Geographic distribution among clusters

In contrast to the greater panel, information regarding origin is available for the majority of k3 accessions (Table 1). All accessions in the k3.3 cluster were collected in Zambia indicating this grouping may be native to south-central Sub-Sahara Africa. Accessions in k3.1 and k3.2 are far less geographically congruent. Sub-cluster k3.3 includes diverse New and Old World geographic collections ranging from Uruguay (182; PI 511865) to Sri Lanka (179; PI 652055) to Tanzania (178; PI 652064). The collection locations of four accessions 116 - 118 in k2.1 are split between Turkey (PI 172996; PI 172998) and Iran (PI 253157; 296391) suggesting a Middle-Eastern grouping. The remaining k2.1 accessions are cultivars with unknown parentage developed for a volatile composition high in monoterpenoids geranial and neral (i.e., citral). The k2.2 sub-cluster includes 123 (PI 652060) and 122 (PI 652061) collected in Pakistan and India, respectively, however, geographic origin of the remaining five accessions are unknown. The k1 cluster is comprised of largely commercial and Rutgers breeding lines in which parentage or geographical origin of parents is unknown. Ten of 12 k1 USDA-GRIN accessions were reported as collected in Turkey, Iran or Macedonia.

Distribution of DM resistance

Two unique sources of DM resistance (mean DS < 1.0), 47 (‘MRI’) and 83 (‘Kivumbisi Lime x SB17 F1’), were identified in the major cluster k1. The former accession has been used to develop Rutgers DM_{MRI} resistant breeding lines, while the latter is a sterile F_1 hybrid with DM resistance conferred from the admixed accession 153 (‘Kivumbisi Lime’). Apart from these two sources of resistance, the k1 cluster was largely DM susceptible with a mean DS of 3.6. Seven accessions demonstrated an intermediate response to DM (1.0 < DS < 3.0) which could be incorporated into breeding strategies targeting durable resistance [26].

The frequency of DM intermediate and resistant accessions increased with genetic distance from the k1 cluster. Intermediate resistance was observed among the majority of k2 accessions reflected by the 2.27 overall mean DS for this cluster. Interestingly, the k2.1 mean DS of 1.5 was lower than for k2.2 (mean DS = 3.4) for which the lowest DS was 2.9 (Fig. 4). Major cluster k3 is the primary source of DM resistance (mean DS = 0.36) with 29 resistant, 4 intermediate and one susceptible genotype. The highly supported (bootstrap = 0.867) clade of four accessions 156 (‘Holy Basil’) – 159 (PI 652057) exhibited hypersensitive response (HR). Although HR is considered a defense strategy against obligate pathogens, was not completely effective in preventing sporulation for these accessions such as 159 (PI 652057) (DS = 2.57) (Table 1).

Fourteen F_1 hybrids developed from DM resistant parents demonstrated resistance with the exception of accession 100 (‘SB17 x 172996 F1’) with DS = 2.75. These results are in agreement with previous evaluation of F_1 progeny providing corroborating evidence of dominant gene action [43]. Three k1 x k3 DM resistant F_1 hybrids developed from 139 (‘Spice’), 153 (‘Kivumbisi Lime’) and 149 (‘Camphor’) exhibit resistance but

were completely sterile. The k2.1 DM resistant 116 (PI 172996) and intermediate 115 ('Sweet Dani') accessions were also used to develop F1 progeny 100 ('SB17 x PI 172996 F1') and 105 ('Sweet Dani x RUCB_10 F1'), respectively. The intermediate response (DS = 2.25) observed for 115 was not conferred in the F1 accession 105 (DM = 3.38), while F1 accession 100 was intermediate (DS = 2.75) relative to the resistant parent 116. Both F1 progeny demonstrated partial sterility, but fertility was restored in some BC1 progeny providing a viable breeding platform.

Discussion

Population structure and genetic diversity

Model-based and UPGMA clustering analyses provide evidence of three distinct populations and 15.6% admixture among the 180 *Ocimum* spp. accession panel. Nested population structure analysis provides additional evidence for seven sub-populations. Greatest support (ΔK) was provided for two distinct populations (K=2) nested within the most economically important *O. basilicum* k1 cluster. However, the possibility of five and seven sub-populations could not be discounted (Fig. 1B). Cryptic population structure in the k1 cluster may be attributed to extensive interbreeding, which is particularly prevalent in the *O. basilicum* species [49]. High levels of admixture were also observed for the K=5 and K=7 population models (Fig. 1B), further supporting the exchange of genomic content potentially as a function of natural and purposeful outcrossing. This is most evident in the k1.2 cluster, where both admixture (Fig. 1B) and genetic distance increase (Fig. 3B) proceeding towards the basal end containing a large proportion of cultivars and USDA-GRIN accessions (Table 1). These results indicate

plant breeding and natural outcrossing have maintained some level of genetic diversity within this economically salient cluster of accessions.

The k1 cluster was the largest in this study with 90 accessions, yet best evidence was provided for only two sub-clusters, which was also the case for the k2 cluster of just 16 accessions (Table 3). Furthermore, the 36-accession k3 cluster was resolved into three sub-clusters and may potentially contain additional populations based on interpretation of UPGMA analysis (Fig. 3C-D). While these results do not necessarily suggest a genetic bottleneck in k1, the k2 and k3 clusters clearly exhibit far greater genetic diversity. Many accessions in the k1 cluster have been subjected to selection for important horticultural traits (i.e., chilling tolerance and disease resistance). Although this has provided elite breeding lines it may also have reduced overall genetic diversity among this *O. basilicum* germplasm. Introgression of accessions from the k2 and k3 clusters would be advantageous in broadening the k1 (*O. basilicum*) gene pool.

Taxonomical discrepancies

Results of this study demonstrated that genetic distance and population inference was often not correlated with reported accession collection location, phenotype or even species epithet. Conflicting and redundant nomenclature in literature, the USDA-GRIN and commercial seed sources have confounded accurate species-level assignment to important *Ocimum* spp. accessions [24]. Phenotype appears to be somewhat predictive of genotype in clusters k3.1 and k3.2, however, leaf shape, habit, flower morphology and volatiles composition are clearly heterogeneous among the remaining sub-clusters. For instance, k2.1 accessions 115 (‘Sweet Dani’) and 116 (PI 172996) group within a single clade (0.553), however, 115 exhibits no anthocyanins and a volatile composition that is

68% citral [8], while 116 exhibits leaf and stem anthocyanins and a volatile composition 91% methyl chavicol as reported by the USDA-GRIN. Meanwhile, k1.2 cluster 77 (‘Queenette’) and 78 (‘Sweet Thai’) contain stem anthocyanins and high methyl chavicol, comparable to accession 116. Although phenotype-based classification may provide some predictive measure of genetic relatedness, applications such as plant breeding require more precise, highly accurate understanding of relations among accessions in order to construct effective selection strategies.

Ploidy and reproductive barriers

Interesting differences in the mean and distribution of alleles per locus were observed among major and sub-clusters in this study. The majority of non-admixed accessions included in the panel are considered highly inbred and therefore homozygous across loci. Given this general assumption, some inferences can be made with regard ploidy. A majority bi-allelic loci among k1 accessions suggests this cluster is an allotetraploid with two homozygous loci representing two sub-genomes. This system has been previously observed for the well-characterized allotetraploid AACC and A_tD_t genomes of *Brassica napus* [33] and *Gossypium hirsutum* [50], respectively. One and two alleles per locus were observed in 33 and 57% of loci in the k1 cluster. These results are similar to the distribution of alleles per loci in a 240 EST-SSR marker survey using inbred k1 accessions 21 (‘RU_SB22’) and 47 (‘MRI’), in which a polymorphic subset also demonstrated disomic inheritance [17].

Among accessions in the k2 cluster, an 36% rate of three alleles per locus provides evidence of three homeologous loci, suggesting an allohexaploid genome. Previous investigations of cytology and nuclei acid content for the species *O. africanum*

(syn. *O. citriodorum*) [20] suggested an allohexaploid (2n = 6x = 72) genome structure [11,13,18,19]. Allelic distribution and cluster analysis in this study both support the hypothesis of an allohexaploid *O. africanum* k2 cluster and a need for nomenclature to be standardized across these accessions (Table 1). Confusingly, USDA-GRIN accessions in the strongly supported (0.997) sub-cluster (k3.2) are labeled as *O. xafricanum*, illustrating the continued *Ocimum* spp. taxonomic issues and a need for consensus when assigning species epithet. Hybridization of k1 and k2 accessions was previously characterized as an interspecific crosses with low fertility [11,19]. All k1 x k2 accessions in this study, which include 17 F_1, F_2 and $F_{2:3}$, demonstrated low fertility. This is also in agreement with a previous report of low fertility of the PI 172996 x 'Peri' F_1 hybrid [43]. In this study *O. basilicum* accession 'Peri' is replaced with 20 ('RU_SB17') and used as recurrent parent to generate the BC_1 accessions 89 ('RU172S_303') - 91 ('RU172S_315') (Fig. 3A). Three 'RU_172S' BC1 individuals demonstrate increased fertility relative to the F_1, suggesting restoration of reproductive viability may be correlated with increased k1 membership and providing a platform for development of near isogenic lines (NILs).

A majority (60%) of single allele loci among k3 accessions was nearly two- and three-fold greater than k1 and k2, respectively. One of two scenarios are likely to explain the single-allele EST-SSRs among k3 accessions. First, it is possible that a divergent sub-genome is shared among k3 accessions, but does not have adequate homology for transferability of SSR markers derived from *O. basilicum* EST database. This would suggest a system similar to the triplicated *Brassica* genome in which *B. napus* AACC and *B. juncea* AABB in which only the A genomes are homologous [51]. A second possibility is the complete absence of a sub-genome, which would suggest k3 as a diploid

cluster or at least containing a large number of diploid accessions. A diploid genome hypothesis was previously proposed for k3 *O. gratissimum* and *O. tenuiforum* accessions [11,15], which have consistently lower 2C-values [12,13]. Furthermore, the recently sequenced genome of *O. tenuiflorum* genotype 'CIM Ayu' is described as a diploid (2n = 2x = 16) with a 386 Mbp genome [15], more than ten-fold less than tetraploid *O. basilicum* based on 2C estimates [12,13]. Koroch et al. [13] reported a 2C-value for k3.2 accession 160 (PI 652066) (*O. campechianum*) significantly lower ($p<0.001$) than k3.1 accession 143 (PI 652059) (*O. tenuiflorum*). Accession 182 (PI 511865) was the most basal accession in this study (Fig. 3D), yet is reported to have a 2C-value comparable to 143 (PI 652059) [13]. Thus, comparison of previously reported nucleic acid content with allelic distribution, genetic distance estimates and cluster analyses in this study suggest a complex *Ocimum* spp. genome evolution with multiple chromosomal accumulation and/or deletion events.

Three F_1 progeny of k3.1 accession 139 ('Spice') hybridized with *O. basilicum* k1.1 accessions 22 ('RUSB_09'), 6 ('DiGenova') and 47 ('MRI') form a well-supported (0.938) clade with admixed accessions 135 (PI 414201) and 136 (414203) (Fig. 3D). Sterility among these progeny suggest a major reproductive barrier between commercially important *O. basilicum* k1.1 accessions and this least basal, highly supported (1.00) k3 clade. Ben-Naim et al. [43] reported F_1 sterility among progeny of DM resistant accessions 161 (PI 500945) and 168 (PI 500950) hybridized with *O. basilicum* accession 'Peri' [43]. These accessions are found in the k2.2 sub-cluster suggesting the k1 x k3 F1 sterility barrier extends to this more basal clade (Fig. 3D). Surprising k1 *qI* among a number of non-hybrid k3 accessions 149 ('Camphor') – 155

('African') (Fig. 3D) suggests the possibility of recently shared ancestry between these clusters. However, sterility among 153 ('Kivumbisi Lime') x 20 ('RUSB_17') progeny demonstrates high k1.1 *qI* does not correspond to fertility. Viable F_1 progeny could not be obtained from cross-pollination of k1.1 and k3.2 accessions. This is consistent with at least one previous report of attempted *O. basilicum* x *O. tenuiflorum* hybridization [49].

Although evidence of varying ploidy levels is a suspected cause of reproductive barriers among wide crosses [43] the exact mechanism remains unclear. Comparative genomic analysis of cultivated and wild *Solanum* spp. demonstrate seed and pollen infertility can be controlled by a small number of loci with shared evolutionary history [52]. Further investigation is needed to determine the cause of infertility among *Ocimum* spp.

Distribution of DM resistance and breeding

Forty-three unique sources of genetic resistance to DM are identified from greenhouse screening in this study. These accessions are heavily concentrated in k3 cluster with 29 from seven species according to the sources from which they were attained (Table 1). An additional nine resistant accessions were admixed (*qI*<0.70), but had a majority k3 *qI* and UPGMA placement among k3 accessions (Fig. 3D). It is possible that loci conferring resistance are shared among closely related individuals and therefore the number of unique genes may be lower due to redundancy. This may be the hypothesized scenario for phenotypically indistinguishable accessions of the highly supported (1.00) k3.1 clade (Fig. 3C).

Accessions 116 (PI 172996), 117 (PI 172998) and 119 (PI 296391) were first identified as having little or no disease incidence by Pyne et al. [39]. These accessions

demonstrated minimal leaf sporulation in this study (DS ≤ 0.63) and were determined to be in a population (k2) distinct from commercial sweet basil (k1). MRI represents the only source of DM resistance (DM_{MRI}) identified in the *O. basilicum* k1 cluster. Recently discovered QTL *dm11.1* demonstrated dominant gene action conferred by MRI, but detection of additional minor QTL explained additional phenotypic variance and a somewhat complex mechanism for the highest levels of resistance [17].

These accessions represent candidates for field confirmation of resistance to *P. belbahrii* and, ultimately, use in DM resistance breeding. DM response in greenhouse and field experiments were recently shown to be largely correlated [43]. Response of accessions common across this and previous greenhouse studies [43,44] is generally consistent with minor differences. Differential disease response of genotypes across studies may be due to a number of confounding effects including environment, inoculum concentration and rating system [53]. Another possibility of greater consequence to disease control is the potential occurrence of pathogen evolution resulting in races with differential host virulence. Proliferation of races is common among the most economically important downy mildews such as *Pseudoperonospora cubensis* [54], *Bremia lactuca* [38] and *Hyaloperonospora brassicae* [55]. Phylogenetic characterization of *P. belbahrii* isolates is needed to determine the range of genetic diversity and identify possible extant pathogen races.

Candidacy of accessions for DM resistance breeding is contingent upon sexual compatibility and reproductive capacity with commercial sweet basil found in sub-cluster k1.1 (Fig 3A). DM resistance from k1.1 accession MRI is the only documented characterization [17,56] and introgression of genetic resistance beyond F_1 progeny.

Partial sterility observed for k1 x k2 crosses appears to be surmountable by backcrossing as demonstrated by the 'RU_172S17' BC_1 accessions 89-92 (Fig. 3B). This represents a potential strategy for introduction of k2 genomic DNA associated with important traits such as DM tolerance (Table1) and FOB resistance available in many of these accessions. An intermediate DM response (DS = 2.75) for the SB17 x PI 172996 F1 hybrid was significantly ($P<0.05$) higher than resistant parent PI 172996 and closer to the mid-parent DS of 2.18 suggesting the possibility of quantitative resistance from the k2 cluster. The heritability of DM resistance from k2 accessions such as PI 172996 is unclear and requires further investigation.

The k3 cluster represents a rich source of genetic resistance. Furthermore, consistent non-significant differences in disease response for F_1 progeny and their resistant parent indicate dominant gene action is widespread among these candidates [43]. Qualitative and, in particular, single dominant gene control of DM is well documented in other plant species [38,57,58]. Hypersensitive response in *O. tenuiflorum* accessions 156-159 indicates effector triggered immunity [59], a response distinct from the remaining k3 accessions in which no signs or symptoms were observed. It remains unclear whether these accessions have evolved unique resistance genes, but nonhost resistance is more likely among distant relatives of susceptible k1 accessions. Introgression of nonhost resistance from wild lettuce relatives *Lactuca serriola* and *Lactuca saligna* to commercial *L. sativa* is being used to build durable DM resistance [38,60]. Further investigation is needed to identify redundancy of resistant genes, especially among closely related accessions. More fundamentally, methods such as genome duplication need to be

optimized to overcome reproductive barriers and facilitate introgression between *Ocimum* spp.

Conclusions

Results of this study provide a robust characterization of population structure, genetic diversity and response to DM among *Ocimum* spp. EST-SSRs provide a useful tool for increased resolution of this genus through continued germplasm collection and genotyping. Although evidence is concurrent for allotetraploid and allohexaploid genomes among economically important k1 and k2 clusters, respectively, the highly diverse and DM resistance-rich k3 cluster remains complicated by discrepancies in reports of taxonomy and genome size. Characterization of subgenomes as shared or distinct among allopolyploid species and identification of a diploid progenitor(s) are needed to elucidate genome structure.

As breeding among specialty crops becomes increasingly prevalent more sophisticated tools will be required. This trend is evidenced in *Ocimum* spp. by recent development of a first draft genome [15], *de novo* meta-transcriptomics [61,62] and genetic/QTL mapping [17]. Determination of major and cryptic population structure as well as phylogenetic classification among DM resistant candidates provides another important resource for rapid genetic improvement.

References

1. Paton A, Harley MR, Harley MM. *Ocimum*: An overview of classification and relationships. In: Hiltunen R, Holm Y, editors. Basil: The Genus *Ocimum*. Amsterdam: Harwood Academic Publishers; 1999. pp. 1–38.
2. Vieira RF, Goldsbrough P, Simon JE. Genetic diversity of basil (*Ocimum* spp.) based on RAPD markers. J Am Soc Hort Sci. 2003; 128: 94–99.
3. Paton AJ, Springate D, Suddee S, Otieno D, Grayer RJ, Harley MM, et al. Phylogeny and evolution of basils and allies (*Ocimeae*, *Labiatae*) based on three plastid DNA regions. Mol Phylogenet Evol. 2004; 31: 277–299. doi:10.1016/j.ympev.2003.08.002
4. Simon, J. E.; Quinn, J.; Murray JG. Basil: A Source of Essential Oils. In: Janick, J.; Simon JE, editor. Advances in New Crops. Portland, OR: Timber Press; 1990. pp. 484–489.
5. Putievsky E, Galambosi B. Production systems of sweet basil. In: Hiltunen R, Holm Y, editors. Basil: The Genus *Ocimum*. Amsterdam; 1999. pp. 39–61.
6. Chaimovitsh D, Dudai N, Putievesky E, Ashri A. Inheritance of Resistance to Fusarium Wilt in Sweet Basil. Phytopathology. 2006;90: 58–60. doi:10.2135/cropsci1983.0011183X002300010010x
7. Ribeiro P, Simon JE. Breeding Sweet Basil for Chilling Tolerance. ASHS. 2007; 302–305.
8. Morales MR, Simon JE. 'Sweet Dani': A new culinary and ornamental basil.pdf. HortScience. 1997;1: 148–149.
9. Nation RG, Janick J, Simon JE. Estimation of outcrossing in basil. Hortic Sci. 1992;27: 1221–1222.
10. Khosla MK, Tawi- J, Group S. Karyomorphological Studies in Genus *Ocimum*. Cytologia. 1985;50: 253–263.
11. Pushpangandan P, Sobti SN. Cytogenetical studies in the genus *Ocimum*. I. Origin of *O. americanum*, cytotaxonomical and experimental proof. Cytologia. 1982; 47: 575–583.
12. Carović-Stanko K, Liber Z, Besendorfer V, Javornik B, Bohanec B, Kolak I, et al. Genetic relations among basil taxa (*Ocimum* L.) based on molecular markers, nuclear DNA content, and chromosome number. Plant Syst Evol. 2010;285: 13–22. doi:10.1007/s00606-009-0251-z
13. Koroch AR, Wang W, Michael TP, Dudai N, Simon JE, Belanger FC. Estimation of nuclear DNA content of cultivated *Ocimum* species by using flow cytometry. Isr J Plant Sci. 2010;58: 183–189. doi:10.1560/IJPS.59.3-4.183
14. Escudero M, Martín-Bravo S, Mayrose I, Fernández-Mazuecos M, Fiz-Palacios O, Hipp AL, et al. Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. PLoS One. 2014;9. doi:10.1371/journal.pone.0085266
15. Rastogi S, Kalra A, Gupta V, Khan F, Lal RK, Tripathi AK, et al. Unraveling the genome of Holy basil: an "incomparable" "elixir of life" of traditional Indian medicine. BMC Genomics. 2015;16: 413. doi:10.1186/s12864-015-1640-z
16. Ramsey J, Schemske DW. Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants. Annu Rev Ecol Syst. 1998;29: 467–501.

doi:10.1146/annurev.ecolsys.29.1.467
17. Pyne R, Vaiciunas J, Honig J, Koroch AR, Wyenandt CA, Bonos SA, et al. A first linkage map and downy mildew resistance QTL discovery for sweet basil (*Ocimum basilicum*) facilitated by double digestion restriction site associated DNA sequencing (ddRADseq). PLoS One. 2017; Submitted.
18. Carović-Stanko K, Liber Z, Politeo O, Strikić F, Kolak I, Milos M, et al. Molecular and chemical characterization of the most widespread *Ocimum* species. Plant Syst Evol. 2011;294: 253–262. doi:10.1007/s00606-011-0471-x
19. Paton AJ, Putievesky E. Taxonomic problems and cytotaxonomic relationships between and within varieties of *Ocimum basilicum* and related species (Labiateae). Kew Bull. 1996;51: 509–524.
20. Suddee S, Paton AJ, Parnell JAN. Taxonomic Revision of tribe *Ocimeae* Dumort . (*Lamiaceae*) in continental South East Asia III . *Ociminae*. Kew Bull. 2005; 60: 3-73.
21. Carović-Stanko K, Šalinović A, Grdĭa M, Liber Z, Kolak I, Satovic Z. Efficiency of morphological trait descriptors in discrimination of *Ocimum basilicum* L. accessions. Plant Biosyst. 2011;145: 298–305. doi:10.1080/11263504.2011.558677
22. Grayer RJ, Kite GC, Goldstone FJ, Bryan SE, Paton A, Putievsky E. Infraspecific taxonomy and essential oil chemotypes in sweet basil, *Ocimum basilicum*. Phytochemistry. 1996;43: 1033–1039. doi:10.1016/S0031-9422(96)00429-3
23. Vieira RF, Simon JE. Chemical characterization of basil (*Ocimum* spp.) based on volatile oils. Flavour Fragr J. 2006;21: 214–221. doi:10.1002/ffj.1513
24. Paton A, Harley MR, Harley MM. *Ocimum*: An overview of classification and relationships. In: Hiltunen R, Holm Y, editors. Basil: The Genus *Ocimum*. Amsterdam: Hardwood Academic; 1996. pp. 1–38.
25. Simpson MG. Plant Systematics. 2nd ed. Academic Press; 2010.
26. Fehr WR. Principles of Cultivar Development. New York, NY: MacMillan Publishing Company; 1987.
27. Labra M, Miele M, Ledda B, Grassi F, Mazzei M, Sala F. Morphological characterization, essential oil composition and DNA genotyping of *Ocimum basilicum* L. cultivars. Plant Sci. 2004;167: 725–731. doi:10.1016/j.plantsci.2004.04.026
28. Vieira RF, Grayer RJ, Paton A, Simon JE. Genetic diversity of *Ocimum gratissimum* L. based on volatile oil constituents, flavonoids and RAPD markers. Biochem Syst Ecol. 2001;29: 287–304. doi:10.1016/S0305-1978(00)00062-4
29. Kumar A, Mishra P, Baskaran K, Shukla AK, Shasany AK, Sundaresan V. Higher efficiency of ISSR markers over plastid *psbA-trnH* region in resolving taxonomical status of genus *Ocimum* L. Ecol Evol. 2016;6: 7671–7682. doi:10.1002/ece3.2483
30. Mahajan V, Rather IA, Awasthi P, Anand R, Gairola S, Meena SR, et al. Development of chemical and EST-SSR markers for *Ocimum* genus. Ind Crops Prod. Elsevier B.V.; 2015;63: 65–70. doi:10.1016/j.indcrop.2014.10.052
31. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: An overview of the recent progress in plants. Euphytica. 2011;177: 309–334. doi:10.1007/s10681-010-0286-9
32. Ramchiary N, Nguyen VD, Li X, Hong CP, Dhandapani V, Choi SR, et al. Genic

microsatellite markers in *brassica rapa*: Development, characterization, mapping, and their utility in other cultivated and wild brassica relatives. DNA Res. 2011; 18: 305–320. doi:10.1093/dnares/dsr017

33. Li H, Younas M, Wang X, Li X, Chen L, Zhao B, et al. Development of a core set of single-locus SSR markers for allotetraploid rapeseed (*Brassica napus* L.). Theor Appl Genet. 2013;126: 937–947. doi:10.1007/s00122-012-2027-z
34. Belbahri L, Calmin G, Pawlowski J, Lefort F. Phylogenetic analysis and real time PCR detection of a presumably undescribed *Peronospora* species on sweet basil and sage. Mycol Res. 2005; 109: 1276–1287. doi:10.1017/S0953756205003928
35. Wyenandt CA, Simon JE, Pyne RM, Homa K, McGrath MT, Zhang S, et al. Basil Downy Mildew (*Peronospora belbahrii*): Discoveries and Challenges Relative to Its Control. Phytopathology. 2015;105: 885–94. doi:10.1094/PHYTO-02-15-0032-FI
36. Cohen Y, Van den Langenberg KM, Wehner TC, Ojiambo PS, Hausbeck M, Quesada-Ocampo LM, et al. Resurgence of *Pseudoperonospora cubensis*: The Causal Agent of Cucurbit Downy Mildew. Phytopathology. 2015;105: 998–1012. doi:10.1093/pcp/pci095
37. Call AD, Criswell AD, Wehner TC, Ando K, Grumet R. Resistance of cucumber cultivars to a new strain of cucurbit downy mildew. HortScience. 2012;47: 171–178.
38. Parra L, Maisonneuve B, Lebeda A, Schut J, Christopoulou M, Jeuken M, et al. Rationalization of genes for resistance to *Bremia lact*ucae in lettuce. Euphytica. Springer Netherlands; 2016;210: 309–326. doi:10.1007/s10681-016-1687-1
39. Pyne RM, Koroch AR, Wyenandt CA, Simon JE. A Rapid Screening Approach to Identify Resistance to Basil Downy Mildew (*Peronospora belbahrii*). HortScience. 2014;49: 1041–1045.
40. Cohen Y, Vaknin M, Ben-Naim Y, Rubin AE, Galperin M, Silverman D, et al. First Report of the Occurrence and Resistance to Mefenoxam of *Peronospora belbahrii*, Causal Agent of Downy Mildew of Basil (*Ocimum basilicum*) in Israel. Plant Dis. 2013;97: 692. doi:10.1094/PDIS-12-12-1126-PDN
41. Pintore I, Gilardi G, Gullino ML, Garibaldi A. Detection of mefenoxam-resistant strains of *Peronospora belbahrii*, the causal agent of basil downy mildew, transmitted through infected seeds. Phytoparasitica. 2016; doi:10.1007/s12600-016-0538-x
42. Collina M, Merighi M, Turan C, Pirondi A, Minuto G, Brunelli A. First Report of Resistance of *Peronospora belbahrii*, Causal Agent of Downy Mildew of Basil, to Mefenoxam in Italy. Plant Dis. 2016;100: 1787. doi:10.1094/PDIS-02-16-0237-PDN
43. Ben-Naim Y, Falach L, Cohen Y. Resistance against basil downy mildew in *Ocimum* species. Phytopathology. 2015;105: 778-785. doi:10.1094/PHYTO-11-14-0295-R
44. Djalali Farahani-Kofoet R, Romer P, Grosch R. Selecting basil genotypes with resistance against downy mildew. Sci Hortic. 2014;179: 248–255. doi:10.1016/j.scienta.2014.09.036
45. Wyenandt CA, Simon JE, McGrath MT, Ward DL. Susceptibility of basil cultivars and breeding lines to downy mildew (*Peronospora belbahrii*). HortScience. 2010;

45: 1416–1419.
46. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. 2000; 155: 945-955.
47. Zhu C, Gore M, Buckler ES, Yu J. Status and Prospects of Association Mapping in Plants. Plant Genome. 2008;1: 5-20. doi:10.3835/plantgenome2008.02.0089
48. Rohlf FJ. NTSYSpc: Numerical Taxonomy System. Setauket, NY: Exeter Publishing, Ltd.; 2008.
49. Putievesky E, Paton A, Lewinsohn E, Ravid U, Haimovich D, Katzir I, et al. Crossability and relationship between *Ocimum basilicum* L. J Herbs, Spices, Med Plants. 1999; 6: 11–24.
50. Zhao Y, Wang H, Chen W, Li Y. Genetic structure, linkage disequilibrium and association mapping of verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. PLoS One. 2014;9. doi:10.1371/journal.pone.0086308
51. Cheng F, Wu J, Wang X. Genome triplication drove the diversification of *Brassica* plants. 2014; 1–8. doi:10.1038/hortres.2014.24
52. Moyle LC, Nakazato T. Comparative Genetics of Hybrid Incompatibility: Sterility in Two *Solanum* Species Crosses. 2008;1453: 1437–1453. doi:10.1534/genetics.107.083618
53. Agrios GN. Plant Pathology. 5th ed. Burlington, MA: Elsevier Academic Press; 2005.
54. Ojiambo PS, Gent DH, Quesada-Ocampo LM, Hausbeck MK, Holmes GJ. Epidemiology and Population Biology of *Pseudoperonospora cubensis* : A Model System for Management of Downy Mildews. Annu Rev Phytopathol. 2015;53: 223–246. doi:10.1146/annurev-phyto-080614-120048
55. Vicente JG, Gunn ND, Bailey L, Pink DAC, Holub EB. Genetics of resistance to downy mildew in *Brassica oleracea* and breeding towards durable disease control for UK vegetable production. Plant Pathol. 2012;61: 600–609. doi:10.1111/j.1365-3059.2011.02539.x
56. Pyne RM, Koroch AR, Wyenandt CA, Simon JE. Inheritance of Resistance to Downy Mildew in Sweet Basil. J Am Soc Hortic Sci. 2015;140: 396–403.
57. Irish BM, Correll JC, Feng C, Bentley T, de Los Reyes BG. Characterization of a resistance locus (*Pfs-1*) to the spinach downy mildew pathogen (*Peronospora farinosa* f. sp. *spinaciae*) and development of a molecular marker linked to *Pfs-1*. Phytopathology. 2008;98: 894–900. doi:10.1094/PHYTO-98-8-0894
58. Epinat C, Pitrat M. Inheritance of resistance of three lines of muskmelon (*Cucumis melo*) to downy mildew (*Pseudoperonospora cubensis*). Proc Cucurbitaceae: Evaluation and Enhancement of Cucurbit Germplasm. Charleston, SC; 1989. pp. 133–135.
59. Jones JDG, Dangl L. The plant immune system. Nature. 2006;444: 323–329. doi:10.1038/nature05286
60. Zhang NW, Lindhout P, Niks RE, Jeuken MJW. Genetic dissection of *Lactuca saligna* nonhost resistance to downy mildew at various lettuce developmental stages. Plant Pathol. 2009;58: 923–932. doi:10.1111/j.1365-3059.2009.02066.x
61. Guo L, Allen KS, Deiulio GA, Zhang Y, Madeiras AM, Wick RL, et al. A de-novo-assembly-based Data Analysis Pipeline for Plant Obligate Parasite

Metatranscriptomic Studies. Front Plant Sci. 2016;7: 925. doi:10.3389/FPLS.2016.00925
62. Rastogi S, Meena S, Bhattacharya A, Ghosh S, Shukla RK, Sangwan NS, et al. De novo sequencing and comparative analysis of holy and sweet basil transcriptomes. BMC Genomics. 2014; 15: 588. doi:10.1186/1471-2164-15-588
63. van Dijk T, Noordijk Y, Dubos T, Bink MC, Meulenbroek BJ, Visser RG, et al. Microsatellite allele dose and configuration establishment (MADCE): an integrated approach for genetic studies in allopolyploids. BMC Plant Biol. 2012;12: 25. doi:10.1186/1471-2229-12-25
64. Honig JA, Kubik C, Averello V, Vaiciunas J, Meyer WA, Bonos SA. Classification of bentgrass (*Agrostis*) cultivars and accessions based on microsatellite (SSR) markers. Genet Resour Crop Evol. 2016; 63: 1139–1160. doi:10.1007/s10722-015-0307-6
65. Vähä JP, Erkinaro J, Niemelä E, Primmer CR. Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. Mol Ecol. 2007; 16: 2638–2654. doi:10.1111/j.1365-294X.2007.03329.x
66. Urrestarazu J, Denancé C, Ravon E, Guyader A, Guisnel R, Feugey L, et al. Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. BMC Plant Biol. 2016; 16: 1–20. doi:10.1186/s12870-016-0818-0
67. Nachimuthu VV, Muthurajan R, Duraialaguraja S, Sivakami R, Pandian BA, Ponniah G, et al. Analysis of Population Structure and Genetic Diversity in Rice Germplasm Using SSR Markers: An Initiative Towards Association Mapping of Agronomic Traits in *Oryza Sativa*. Rice; 2015; 8: 30. doi:10.1186/s12284-015-0062-5
68. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE : a simulation study. Mol Ecol. 2005; 14: 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
69. Earl DA, Bridgett M. STRUCTURE HARVESTER : a website and program for visualizing STRUCTURE output and implementing the Evanno method. 2012; 359–361. doi:10.1007/s12686-011-9548-7
70. Jakobsson M, Rosenberg NA. CLUMPP : a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. 2007; 23: 1801–1806. doi:10.1093/bioinformatics/btm233
71. Rosenberg NA. DISTRUCT : a program for the graphical display of population structure. 2004;4: 137–138. doi:10.1046/j.1471-8286.2003.00566.x
72. Taranto F, D'Agostino N, Greco B, Cardi T, Tripodi P. Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. BMC Genomics. BMC Genomics; 2016;17: 943. doi:10.1186/s12864-016-3297-7
73. Peakall R, Smouse PE. GenAlEx 6.5 : genetic analysis in Excel. Population genetic software for teaching and research - an update. 2012; 28: 2537–2539. doi:10.1093/bioinformatics/bts460
74. Nei M. Genetic Distance between Populations. Am Soc Nat. 2010;106: 283–292.
75. Jaccard P. Nouvelles recherches sur la distribution florale. Bull la Soc Vaudoise des Sci Nat. 1908;44: 223–270.

Table 1. Description of 180-accession panel of *Ocimum* spp., cluster membership and response to downy mildew (*P. belbahrii*) measured as disease severity (DS)

Entry ID	Origin/Source[a]	Accession	Species[b]	Cluster[c]	Sub-Cluster[c]	DS[e]	
3	HSC	Caesar	*O. basilicum*	k1	k1.1	4.00	a
4	Rutgers	RUSB_23	*O. basilicum*	k1	k1.1	3.75	abc
5	EZ	Genovese Martina	*O. basilicum*	k1	k1.1	4.00	a
6	SSC	DiGenova	*O. basilicum*	k1	k1.1	4.00	a
7	EZ	Eowyn	*O. basilicum*	k1	k1.1	4.00	a
8	Rutgers	RUCB_01	*O. basilicum*	k1	k1.1	3.88	ab
9	Rutgers	RUCB_39	*O. basilicum*	k1	k1.1	3.50	abcde
10	Rutgers	RUCB_04	*O. basilicum*	k1	k1.1	3.75	abc
11	Rutgers	RUCB_10	*O. basilicum*	k1	k1.1	4.00	a
12	Rutgers	RUCB_31	*O. basilicum*	k1	k1.1	4.00	a
13	Rutgers	RUSB_13	*O. basilicum*	k1	k1.1	2.38	fghij
14	Rutgers	RUCB_16	*O. basilicum*	k1	k1.1	3.13	abcdef
15	Rutgers	RUCB_14	*O. basilicum*	k1	k1.1	3.63	abcd
16	Rutgers	RUCB_19	*O. basilicum*	k1	k1.1	3.75	abc
17	Rutgers	Poppy Joes	*O. basilicum*	k1	k1.1	4.00	a
18	JSS	Italian Large Leaf	*O. basilicum*	k1	Admixed	4.00	a
19	JSS	Nufar	*O. basilicum*	k1	k1.1	4.00	a
20	Rutgers	RUSB_17	*O. basilicum*	k1	k1.1	3.63	abcd
21	Rutgers	RUSB_22	*O. basilicum*	k1	k1.1	4.00	a

22	Rutgers	RUSB_09	*O. basilicum*	k1	k1.1	4.00	a
23	Rutgers	RUSB_05	*O. basilicum*	k1	k1.2	3.88	ab
24	Greece	PI 263870	*O. basilicum*	k1	k1.1	3.88	ab
25	RSS	Marseilles	*O. basilicum*	k1	k1.2	3.63	abcd
26	Maryland	PI 414197	*O. basilicum*	k1	k1.2	4.00	a
27	Rutgers	RUSB_06	*O. basilicum*	k1	Admixed	4.00	a
28	Rutgers	RU_003_4	*O. basilicum x O. americanum*	k1	Admixed	2.75	defgh
29	Hawaii	Haw1	*O. basilicum*	k1	Admixed	4.00	a
30	JSS	Napoletano	*O. basilicum*	k1	Admixed	4.00	a
31	HSC	Lettuce Leaf Heirloom	*O. basilicum*	k1	Admixed	4.00	a
32	Rutgers	MRI x RUSB_22 F1	*O. basilicum*	k1	k1.1	0.88	nopqrs
33	Rutgers	RUMS_707	*O. basilicum*	k1	k1.1	0.75	opqrs
34	Rutgers	RUMS_707010103	*O. basilicum*	k1	k1.1	0.75	opqrs
35	Rutgers	RUMS_569	*O. basilicum*	k1	k1.1	0.38	qrs
36	Rutgers	RUMS_46911	*O. basilicum*	k1	k1.1	0.63	pqrs
37	Rutgers	RUMS_498	*O. basilicum*	k1	k1.1	4.00	a
38	Rutgers	RUMS_469	*O. basilicum*	k1	k1.1	0.56	qrs
39	Rutgers	RUMS_394	*O. basilicum*	k1	k1.1	0.63	pqrs
40	Rutgers	RU4S_33	*O. basilicum*	k1	k1.1	0.25	rs
41	Rutgers	RU4S_36	*O. basilicum*	k1	k1.1	0.13	rs
42	Rutgers	RU4S_37	*O. basilicum*	k1	k1.1	0.13	rs
43	Rutgers	RU4S_50	*O. basilicum*	k1	k1.1	0.38	qrs
44	Rutgers	RU4S_0177	*O. basilicum*	k1	k1.1	0.50	qrs

44	Rutgers	RU4S_0177	*O. basilicum*	k1	k1.1	0.50	qrs
45	Rutgers	RUMS_394600115	*O. basilicum*	k1	k1.1	0.63	pqrs
46	Rutgers	RU4S_0741	*O. basilicum*	k1	k1.1	0.38	qrs
47	SMS	MRI	*O. basilicum*	k1	k1.1	0.25	rs
48	Rutgers	RU4S_07410401	*O. basilicum*	k1	k1.1	0.13	rs
49	Rutgers	RU4S_07410402	*O. basilicum*	k1	k1.1	0.13	rs
50	Rutgers	RU4S_07410403	*O. basilicum*	k1	k1.1	0.13	rs
51	Rutgers	RU4S_42210101	*O. basilicum*	k1	k1.1	0.25	rs
52	Rutgers	RU4S_42210102	*O. basilicum*	k1	k1.1	0.25	rs
53	Rutgers	RU4S_42210103	*O. basilicum*	k1	k1.1	0.25	rs
54	Rutgers	RU4S_26243301	*O. basilicum*	k1	k1.1	0.13	rs
55	Rutgers	RU4S_26243302	*O. basilicum*	k1	k1.1	0.13	rs
56	Rutgers	RU4S_26243303	*O. basilicum*	k1	k1.1	0.13	rs
57	Rutgers	RU4S_47082301	*O. basilicum*	k1	k1.1	0.75	opqrs
58	Rutgers	RU4S_47082302	*O. basilicum*	k1	k1.1	0.75	opqrs
59	Rutgers	RU4S_47082303	*O. basilicum*	k1	k1.1	0.75	opqrs
60	Ethiopia	PI 197442	*O. basilicum*	k1	k1.2	1.88	hijklm
61	RSS	Globette	*O. basilicum*	k1	Admixed	2.88	cdefg
62	RSS	Minette	*O. basilicum*	k1	Admixed	3.50	abcde
63	SOC	Cinnamon2	*O. basilicum*	k1	k1.2	4.00	a
64	SSC	Red Genovese	*O. basilicum*	k1	k1.2	3.63	abcd
65	Turkey	PI 170581	*O. basilicum*	k1	k1.2	4.00	a
66	Rutgers	RUMS_394601241101	*O. basilicum*	k1	k1.1	0.25	rs
67	Rutgers	RUMS_394601241102	*O. basilicum*	k1	k1.1	0.25	rs

67	Rutgers	RUMS_394601241102	*O. basilicum*	k1	k1.1	0.25	rs
68	Rutgers	RUMS_394601241103	*O. basilicum*	k1	k1.1	0.25	rs
69	EZ	Eleonora	*O. basilicum*	k1	Admixed	1.63	jklmno
70	Turkey	PI 175793	*O. basilicum*	k1	k1.2	4.00	a
71	Iran	PI 190100	*O. basilicum*	k1	k1.2	3.88	ab
72	Macedonia	PI 368697	*O. basilicum*	k1	k1.2	3.38	abcde
73	Macedonia	PI 358469	*O. basilicum*	k1	k1.2	4.00	a
74	RSS	Magical Michael	*O. basilicum*	k1	k1.2	4.00	a
75	Macedonia	PI 358464	*O. basilicum*	k1	k1.2	4.00	a
76	Macedonia	PI 368700	*O. basilicum*	k1	k1.2	4.00	a
77	RSS	Queenette	*O. basilicum*	k1	Admixed	4.00	a
78	JSS	Sweet Thai	*O. basilicum*	k1	k1.2	4.00	a
79	RSS	Thai Siam Queen	*O. basilicum*	k1	k1.2	4.00	a
80	KSC	Siam Queen OP	*O. basilicum*	k1	k1.2	4.00	a
81	Rutgers	RUMC_09	*O. basilicum*	k1	Admixed	4.00	a
82	Macedonia	PI 358465	*O. basilicum*	k1	k1.2	3.88	ab
83	Rutgers	Kivumbisi Lime x RUSB_17 F1	*O. americanum x O. basilicum*	k1	k1.2	0.00	s
84	Turkey	PI 170579	*O. basilicum*	k1	k1.2	3.63	abcd
85	Rutgers	Pesto Perpetuo	*O. basilicum*	k1	k1.2	3.13	abcdef
86	Rutgers	GCB	*O. basilicum*	k1	k1.2	1.38	klmnop
87	Rutgers	Pezou	*O. basilicum*	k1	k1.2	3.88	ab
88	JSS	Cinnamon1	*O. basilicum*	k1	k1.2	4.00	a
93	Turkey	PI 182246	*O. basilicum*	k1	k1.2	4.00	a
94	FS	Violetto Aromatico	*O. basilicum*	k1	k1.2	4.00	a

94	FS	Violetto Aromatico	*O. basilicum*	k1	k1.2	4.00	a
95	JSS	Purple Ruffles	*O. basilicum*	k1	k1.2	1.75	ijklmn
96	JSS	Red Rubin	*O. basilicum*	k1	k1.2	2.13	ghijkl
114	Rutgers	PI 172996 x Sweet Dani F1	*O. basilicum x O. citriodorum*	k2	k2.1	1.75	ijklmn
115	PA	Sweet Dani	*O. citriodorum*	k2	k2.1	2.25	fghijk
116	Turkey	PI 172996	*O. basilicum*	k2	k2.1	0.63	pqrs
117	Turkey	PI 172998	*O. basilicum*	k2	k2.1	0.38	qrs
118	Iran	PI 253157	*O. basilicum*	k2	k2.1	1.88	hijklm
119	Iran	PI 296391	*O. basilicum*	k2	k2.1	0.63	pqrs
120	New Mexico	PI 652054	*O. basilicum*	k2	k2.1	2.13	ghijkl
121	EZ	Lemona	*O. citriodorum*	k2	k2.1	2.13	ghijkl
122	India	PI 652061	*O. basilicum*	k2	k2.2	2.88	cdefg
123	Pakistan	PI 652060	*O. americanum*	k2	k2.2	4.00	a
124	TSC	Lemon	*O. citriodorum*	k2	k2.2	3.13	abcdef
125	BCH	Thai Hairy Lemon	*O. citriodorum*	k2	k2.2	4.00	a
126	KSC	Penang Lemon	*O. citriodorum*	k2	k2.2	3.00	bcdefg
127	BCH	Lime1	*O. americanum*	k2	k2.2	3.63	abcd
128	JSS	Lime2	*O. americanum*	k2	k2.2	3.13	abcdef
129	SOC	Lemon Heirloom	*O. citriodorum*	k2	Admixed	0.88	nopqrs
138	Rutgers	Blue Spice F1	*O. americanum*	k3	k3.1	0.00	s

139	RSS	Spice	*O. americanum*	k3	k3.1	0.00	s
140	Maryland	PI 414205	*O. tenuiflorum*	k3	k3.1	0.00	s
141	Maryland	PI 414204	*O. tenuiflorum*	k3	k3.1	0.00	s
142	Denmark	PI 652056	*O. tenuiflorum*	k3	k3.1	0.00	s
142	Denmark	PI 652056	*O. tenuiflorum*	k3	k3.1	0.00	s
143	Maldives	PI 652059	*O. tenuiflorum*	k3	k3.1	0.00	s
144	Rutgers	Blue Spice	*O. americanum*	k3	k3.1	0.00	s
145	Maryland	PI 652052	*O. x africanum*	k3	k3.2	1.75	ijklmn
149	WSHF	Camphor	*O. kilimandsharicum*	k3	k3.2	0.13	rs
150	Maryland	PI 652053	*O. basilicum*	k3	k3.2	0.25	rs
156	RSS	Holy Basil	*O. tenuiflorum*	k3	k3.2	1.00	mnopqr
157	India	PI 288779	*O. tenuiflorum*	k3	k3.2	0.00	s
158	RSS	Indian	*O. tenuiflorum*	k3	k3.2	0.13	rs
160	Brazil	PI 652066	*O. campechianum*	k3	k3.2	0.00	s
161	Zambia	PI 500945	*O. americanum*	k3	k3.3	0.00	s
162	Zambia	PI 500946	*Ocimum spp.*	k3	k3.3	1.00	mnopqr
163	Zambia	PI 500949	*O. x africanum*	k3	k3.3	0.00	s
164	Zambia	PI 500943	*O. x africanum*	k3	k3.3	1.88	s
165	Zambia	PI 500944	*O. x africanum*	k3	k3.3	0.00	s
166	Zambia	PI 500947	*O. x africanum*	k3	k3.3	1.13	s
167	Zambia	PI 500953	*O. x africanum*	k3	k3.3	0.00	s
168	Zambia	PI 500950	*O. x africanum*	k3	k3.3	0.25	rs
169	Zambia	PI 500954	*O. x africanum*	k3	k3.3	0.00	s

169	Zambia	PI 500954	*O. x africanum*	k3	k3.3	0.00	s
170	Zambia	ZA1	*Ocimum spp.*	k3	k3.3	4.00	a
171	Zambia	PI 500942	*O. x africanum*	k3	k3.3	0.25	s
172	Brazil	PI 652067	*O. gratissimum*	k3	k3.2	0.00	s
173	Brazil	PI 652068	*O. gratissimum*	k3	k3.2	0.00	s
174	Brazil	PI 652069	*O. gratissimum*	k3	k3.2	0.00	s
175	Zambia	PI 500952	*O. gratissimum var. gratissimum*	k3	k3.2	0.00	s
176	Taiwan	PI 211715	*O. gratissimum var. gratissimum*	k3	k3.2	0.00	s
177	SMS	Mtule	*Ocimum spp.*	k3	k3.2	0.00	s
178	Tanzania	PI 652064	*O. gratissimum var. gratissimum*	k3	k3.2	0.25	rs
179	Sri Lanka	PI 652055	*O. gratissimum var. macrophyllum*	k3	k3.2	0.00	s
180	SMS	Mzambda	*Ocimum spp.*	k3	k3.2	0.00	s
181	WSHF	Green Pepper Basil	*O. selloi*	k3	k3.2	0.63	pqrs
182	Uruguay	PI 511865	*O. selloi*	k3	k3.2		
89	Rutgers	RU172S_303	*O. basilicum x O. basilicum*	Admixed	Admixed	3.50	abcde
90	Rutgers	RU172S_322	*O. basilicum x O. basilicum*	Admixed	Admixed	3.75	abc
91	Rutgers	RU172S_306	*O. basilicum x O. basilicum*	Admixed	Admixed	3.63	abcd
92	Rutgers	RU172S_315	*O. basilicum x O. basilicum*	Admixed	Admixed	3.50	abcde

97	Rutgers	RUSB_22 x Lime F1	*O. basilicum x O. americanum*	Admixed	Admixed	3.88	ab
98	Rutgers	PI 652060 x RUSB_17 F1	*O. americanum x O. basilicum*	Admixed	Admixed	4.00	a
99	Rutgers	RU172S_01	*O. basilicum x O. basilicum*	Admixed	Admixed	3.13	abcdef
100	Rutgers	SB17 x PI 172996 F1	*O. basilicum x O. basilicum*	Admixed	Admixed	2.75	defgh
101	Rutgers	PI 172996 x MRI F1	*O. basilicum x O. basilicum*	Admixed	Admixed	0.88	nopqrs
102	Rutgers	RU172S_05	*O. basilicum x O. basilicum*	Admixed	Admixed	1.63	jklmno
103	Rutgers	Sweet Dani x RUMC_09 F1	*O. citriodorum x O. basilicum*	Admixed	Admixed	3.38	abcde
104	Rutgers	Sweet Dani x MRI F1	*O. citriodorum x O. basilicum*	Admixed	Admixed	0.75	opqrs
105	Rutgers	Sweet Dani x RUCB 10 F1	*O. citriodorum x O. basilicum*	Admixed	Admixed	3.38	abcde
106	Rutgers	RU172S_03	*O. basilicum x O. basilicum*	Admixed	Admixed	2.63	efghi
107	Rutgers	RU172S_06	*O. basilicum x O. basilicum*	Admixed	Admixed	1.25	lmnopq
108	Rutgers	RU172S_02	*O. basilicum x O. basilicum*	Admixed	Admixed	2.25	fghijk
109	Rutgers	RU172S_04	*O. basilicum x O. basilicum*	Admixed	Admixed	3.75	abc
110	Rutgers	RUSDCB_172 F2	*O. citriodorum x O. basilicum*	Admixed	Admixed	4.00	a
111	Rutgers	RUSDCB_174 F2	*O. citriodorum x O. basilicum*	Admixed	Admixed	3.50	abcde

112	Rutgers	RUSDCB_15101 F3	*O. citriodorum x O. basilicum*	Admixed	Admixed	3.75	abc
113	Rutgers	RUSDCB_151 F2	*O. citriodorum x O. basilicum*	Admixed	Admixed	3.63	abcd
130	Rutgers	Spice x PI 172996 F1	*O. americanum x O. basilicum*	Admixed	Admixed	0.00	s
131	Rutgers	Spice x Sweet Dani F1	*O. americanum x O. citriodorum*	Admixed	Admixed	0.00	s
132	Rutgers	Camphor x PI 172996 F1	*O. kilimandsharicum x O. basilicum*	Admixed	Admixed	0.25	rs
133	Rutgers	MRI x Spice F1	*O. basilicum x O. americanum*	Admixed	Admixed	0.00	s
134	Rutgers	Spice x RUSB_09 F1	*O. americanum x O. basilicum*	Admixed	Admixed	0.00	s
135	Maryland	PI 414201	*O. tenuiflorum*	Admixed	Admixed	0.00	s
136	Maryland	PI 414203	*O. tenuiflorum*	Admixed	Admixed	0.00	s
137	Rutgers	Spice x DiGenova F1	*O. americanum x O. basilicum*	Admixed	Admixed	0.63	pqrs
146	Rutgers	African Blue Basil	*O. basilicum x O. kilimandsharicum*	Admixed	Admixed	0.00	s
147	Rutgers	PI 172996 x Camphor F1	*O. basilicum x O. kilimandsharicum*	Admixed	Admixed	0.31	rs
148	Rutgers	RUSB_17 x Camphor F1	*O. basilicum x O. kilimandsharicum*	Admixed	Admixed	0.25	rs
151	Iran	PI 253158	*O. americanum*	Admixed	Admixed	0.00	s
152	Iraq	PI 254352	*O. americanum*	Admixed	Admixed	0.13	rs
153	SMS	Kivumbisi Lime	*O. americanum*	Admixed	Admixed	0.00	s
154	Togo	PI 652058	*O. americanum*	Admixed	Admixed	0.13	rs
155	SMS	African	*Ocimum spp.*	Admixed	Admixed	0.00	s

159	Cuba	PI 652057	*O. tenuiflorum*	Admixed	Admixed	2.25	fghijk
1	Rutgers	CR3	*Nepeta cataria*		Outgroup		
2	Rutgers	CR9	*Nepeta cataria*		Outgroup		

[a]Origin refers to the collection location reported by the United States Department of Agriculture National Genetic Resources Program (USDA-GRIN). Sources for abbreviations include Baker Creek Heirloom (BCH), Enza Zaden (EZ), Franchi Sementi (FS), Harris Seed Company (HSC), Johnny's Selected Seeds (JSS), Kitasawa Seed Company (KSC), Richter's Seed Company (RSC), Seeds of Change (SOC), Stokes Seed Company (SSC), Strictly Medicinal Seeds (SMS), Territorial Seed Company (TSC) and Well Sweep Herb Farm (WSHF).

[b]Species reported according to the source from which they were obtained

[c]Cluster and sub-cluster refers to the inferred population derived from primary and secondary (nested) model-based cluster analyses, respectively. Admixed accessions refer to those for which $qI < 0.7$ and /or accessions are from a known bi-parental origin of two different primary clusters.

[d]DS refers to disease severity determined by the extent of sporulation on the abaxial leaf surface. Values with shared letters are not significantly different according to Tukey's highly significant difference (HSD) test, $\alpha = 0.05$

Table 2. Description of 20 EST-SSR markers used to classify 180-accession panel of *Ocimum* spp.

Marker ID	Genbank/ Contg ID	Repeat Motif	Alleles	PIC	Product Size	LG	Position (cM)
OBNJR2sg33	DY333933	(AC)16	19	0.01 - 0.45	271 - 299	14	82.7
OBNJR2cn29	Contig1138	(AC)16	11	0.02 - 0.45	246 - 448	13	54.6
OBNJR2sg04	DY343638	(GA)17	27	0.01 - 0.48	268 - 357	13	40.2
OBNJR2sg30	DY336727	(AG)22	20	0.01 - 0.42	241 - 279		
OBNJR3sg124	DY331703	(GCC)6	11	0.01 - 0.40	161 - 327		
OBNJR3sg19	DY343509	(TCA)6	6	0.01 - 0.46	195 - 432	6	18.9
OBNJR3cn298	Contig2510	(CTA)6	9	0.01 - 0.42	298 - 322		
OBNJR3cn359	Contig2911	(GGC)6	11	0.01 - 0.31	161 - 292		
OBNJR3cn362	Contig2969	(TGA)6	7	0.02 - 0.45	225 - 243	3	78.0
OBNJR3sg155	DY325572	(GTT)7	8	0.02 - 0.40	188 - 263		
OBNJR3sg168	DY323726	(GAA)7	11	0.01 - 0.44	297 - 390		
OBNJR3sg113	DY335879	(CCT)7	12	0.01 - 0.32	313 - 392		
OBNJR3cn56	Contig582	(AGG)7	10	0.01 - 0.45	169 - 240	3	84.7
OBNJR3cn74	Contig715	(CAG)7	12	0.01 - 0.43	211 - 244		
OBNJR3sg145	DY328393	(GCT)8	15	0.01 - 0.45	271 - 313		
OBNJR3cn03	Contig100	(CCA)10	15	0.01 - 0.44	198 - 262		
OBNJR3cn210	Contig1890	(AAG)11	14	0.01 - 0.43	266 - 302		
OBNJR3cn240	Contig2142	(ATA)16	24	0.01 - 0.37	254 - 370	19	88.7
OBNJR3sg13	DY344184	(ACA)10	14	0.01 - 0.36	271 - 312		
OBNJR4cn17	Contig2461	(AAAT)5	13	0.01 - 0.41	164 - 365		

Table 3. Summary of allele distribution among clusters resulting from primary and secondary (nested) model-based cluster analyses

Cluster	Sample size	Total alleles	Average alleles per locus
k1	90	3081	1.75
k2	16	709	2.23
k3	36	907	1.34
Admixture	38	1812	2.58
k1.1	51	1760	1.76
k1.2	28	953	1.74
k2.1	8	363	2.27
k2.2	7	301	2.18
k3.1	7	189	1.41
k3.2	18	452	1.33
k3.3	11	266	1.30
Admixture	50	2225	2.49
Overall	180	6509	1.86

Table 4. Analysis of molecular variance (AMOVA) for non-admixed panel of *Ocimum* spp. using 20 EST-SSR markers among clusters resulting from primary and secondary (nested) model-based cluster analyses

Clustering Iteration	Source	df	SS	MS	Estimated σ^2	Molecular σ^2
Primary[a]	Among Pops	2	997.293	498.646	13.0	43%
	Within Pops	139	2367.390	17.032	17.0	57%
	Total	141	3364.683		30.0	100%
Secondary[b]	Among Pops	6	996.930	166.155	9.0	34%
	Within Pops	123	2108.731	17.144	17.1	66%
	Total	129	3105.662		26.1	100%

[a] $\Phi_{PT} = 0.433$; $p<0.001$
[b] $\Phi_{PT} = 0.344$; $p<0.001$

Table 5. Pairwise Φ_{PT} estimates for clusters resulting from the primary model-based cluster analyses

Cluster	k1	k2	k3
k1	0.000		
k2	1.357	0.000	
k3	0.644	0.552	0.000

All other Φ_{PT} estimates were highly significant ($p<0.001$)

Table 6. Pairwise Φ_{PT} estimates for resulting from the secondary model-based cluster analyses

Sub-Cluster	k1.1	k1.2	k2.1	k2.2	k3.1	k3.2	k3.3
k1.1	0.000	-	-	-	-	-	-
k1.2	0.060	0.000	-	-	-	-	-
k2.1	1.338	1.499	0.000	-	-	-	-
k2.2	1.171	1.272	0.044*	0.000	-	-	-
k3.1	0.602	0.637	0.549	0.322*	0.000	-	-
k3.2	0.718	0.683	0.564	0.407	0.000NS	0.000	-
k3.3	0.764	0.723	0.512	0.381	0.036NS	0.022NS	0.000

NS no significantly different $p>0.05$, * $p<0.05$; all other Φ_{PT} estimates were highly significant ($p<0.001$)

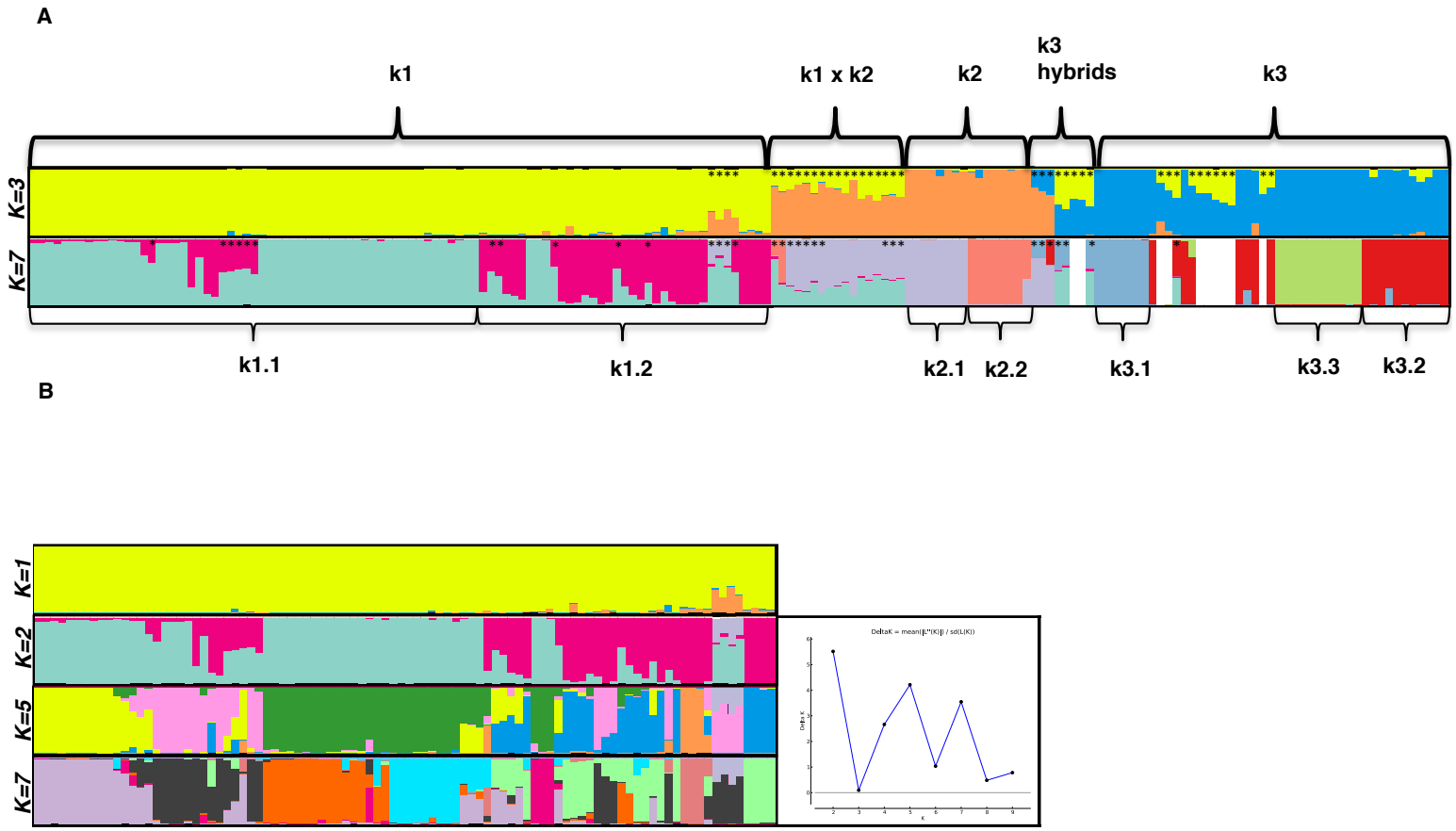


Fig 1. Primary and secondary (nested) model-based clustering analysis using Structure ver 2.2.3 software for panels of *Ocimum* spp. accessions using 20 EST-SSR markers. A. Major clusters (*K*=3) (top histogram) and sub-clusters (*K*=7) (bottom histogram) derived from primary and nested clustering iterations. Ten accessions were admixed and nested population structure could not be inferred due to unknown parentage of admixed primary cluster membership (white bars). B. Major clusters (K=1) (top histogram) and sub-clusters (*K*=2,5,7) (bottom three histograms). ΔK statistic values for K=2-10 (right). Accessions are ordered according to the unweighted pair group method using arithmetic average (UPGMA) clustering

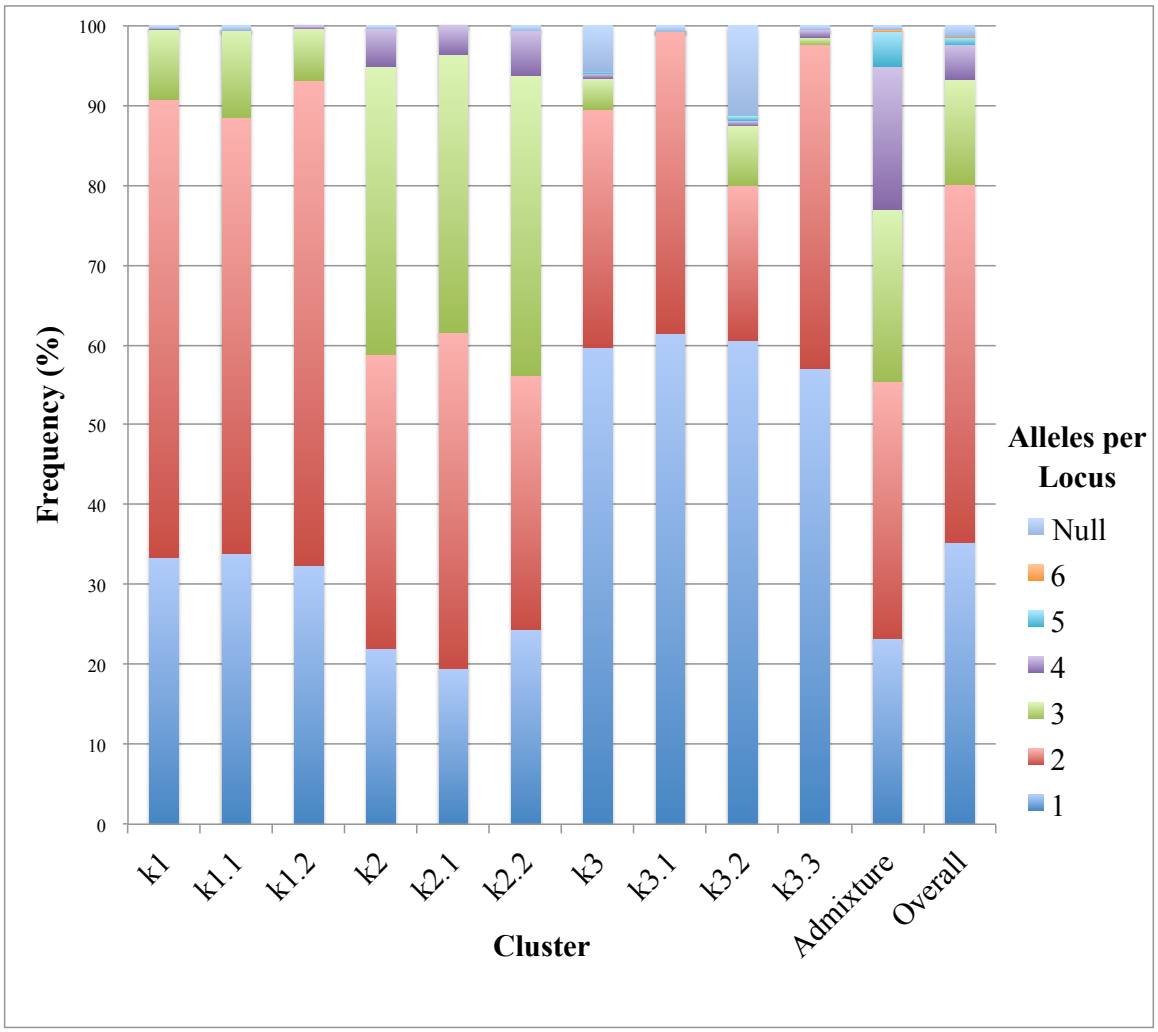
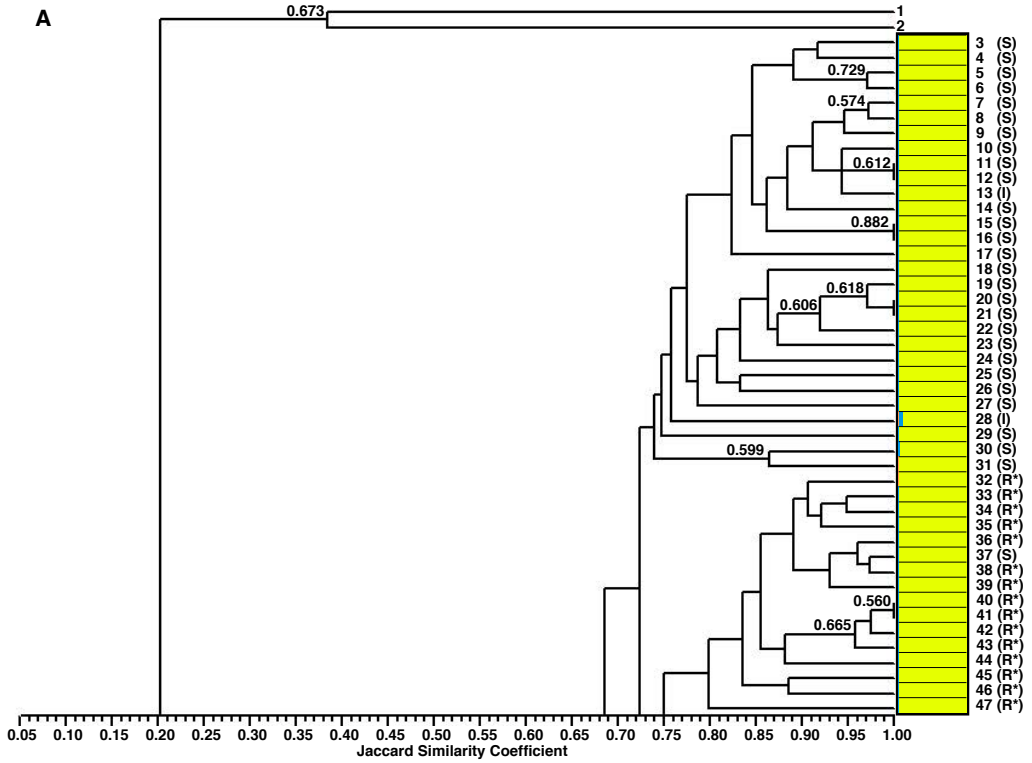
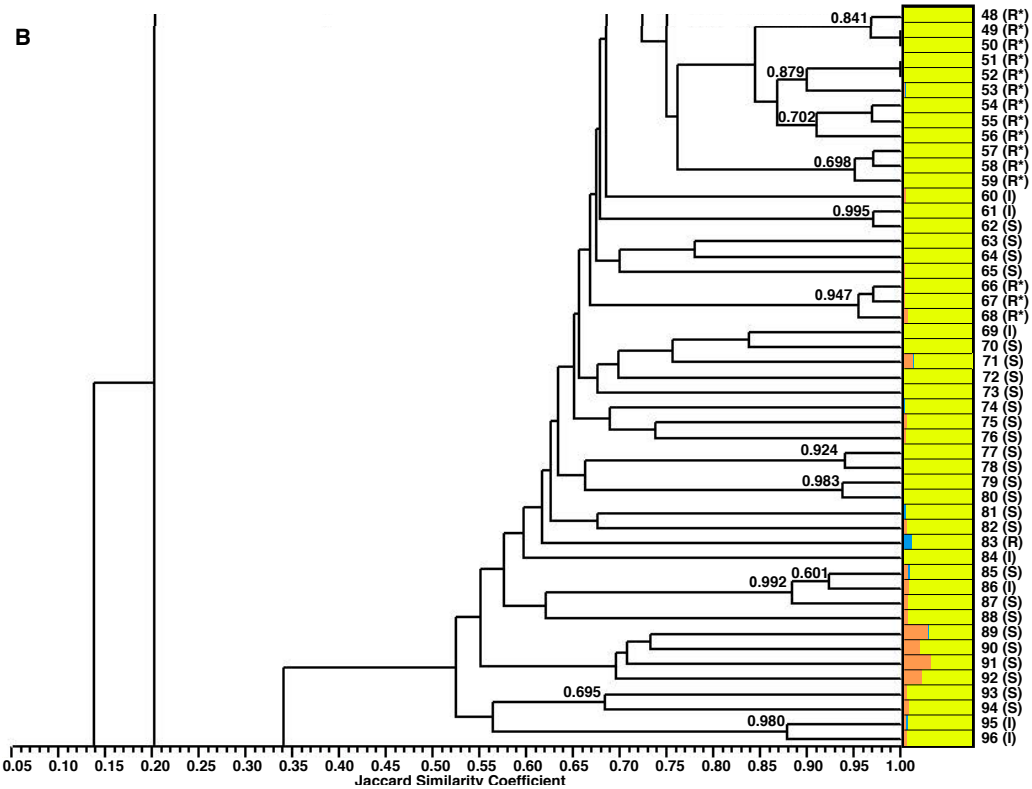


Fig 2. Distribution of alleles per locus for 180-accession panel of *Ocimum* spp. using 20 EST-SSR markers among clusters resulting from primary and secondary (nested) model-based cluster analyses. Colors corresponding to specific numbers of alleles per locus are indicated in the legend (top right). Admixed accessions (n=38) include those with bi-parental parentage between clusters and/or primary clustering *qI* <0.70.


A
0.673
1
2
3 (S)
4 (S)
5 (S)
6 (S)
0.729
7 (S)
0.574
8 (S)
9 (S)
10 (S)
0.612
11 (S)
12 (S)
13 (I)
14 (S)
0.882
15 (S)
16 (S)
17 (S)
18 (S)
19 (S)
0.618
20 (S)
0.606
21 (S)
22 (S)
23 (S)
24 (S)
25 (S)
26 (S)
27 (S)
28 (I)
29 (S)
30 (S)
0.599
31 (S)
32 (R*)
33 (R*)
34 (R*)
35 (R*)
36 (R*)
37 (S)
38 (R*)
39 (R*)
0.560
40 (R*)
41 (R*)
0.665
42 (R*)
43 (R*)
44 (R*)
45 (R*)
46 (R*)
47 (R*)
0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
Jaccard Similarity Coefficient


B
0.841
48 (R*)
49 (R*)
50 (R*)
51 (R*)
0.879
52 (R*)
53 (R*)
54 (R*)
0.702
55 (R*)
56 (R*)
57 (R*)
0.698
58 (R*)
59 (R*)
60 (I)
0.995
61 (I)
62 (S)
63 (S)
64 (S)
65 (S)
66 (R*)
0.947
67 (R*)
68 (R*)
69 (I)
70 (S)
71 (S)
72 (S)
73 (S)
74 (S)
75 (S)
76 (S)
0.924
77 (S)
78 (S)
0.983
79 (S)
80 (S)
81 (S)
82 (S)
83 (R)
84 (I)
0.601
85 (S)
0.992
86 (I)
87 (S)
88 (S)
89 (S)
90 (S)
91 (S)
92 (S)
0.695
93 (S)
94 (S)
0.980
95 (I)
96 (I)
0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
Jaccard Similarity Coefficient

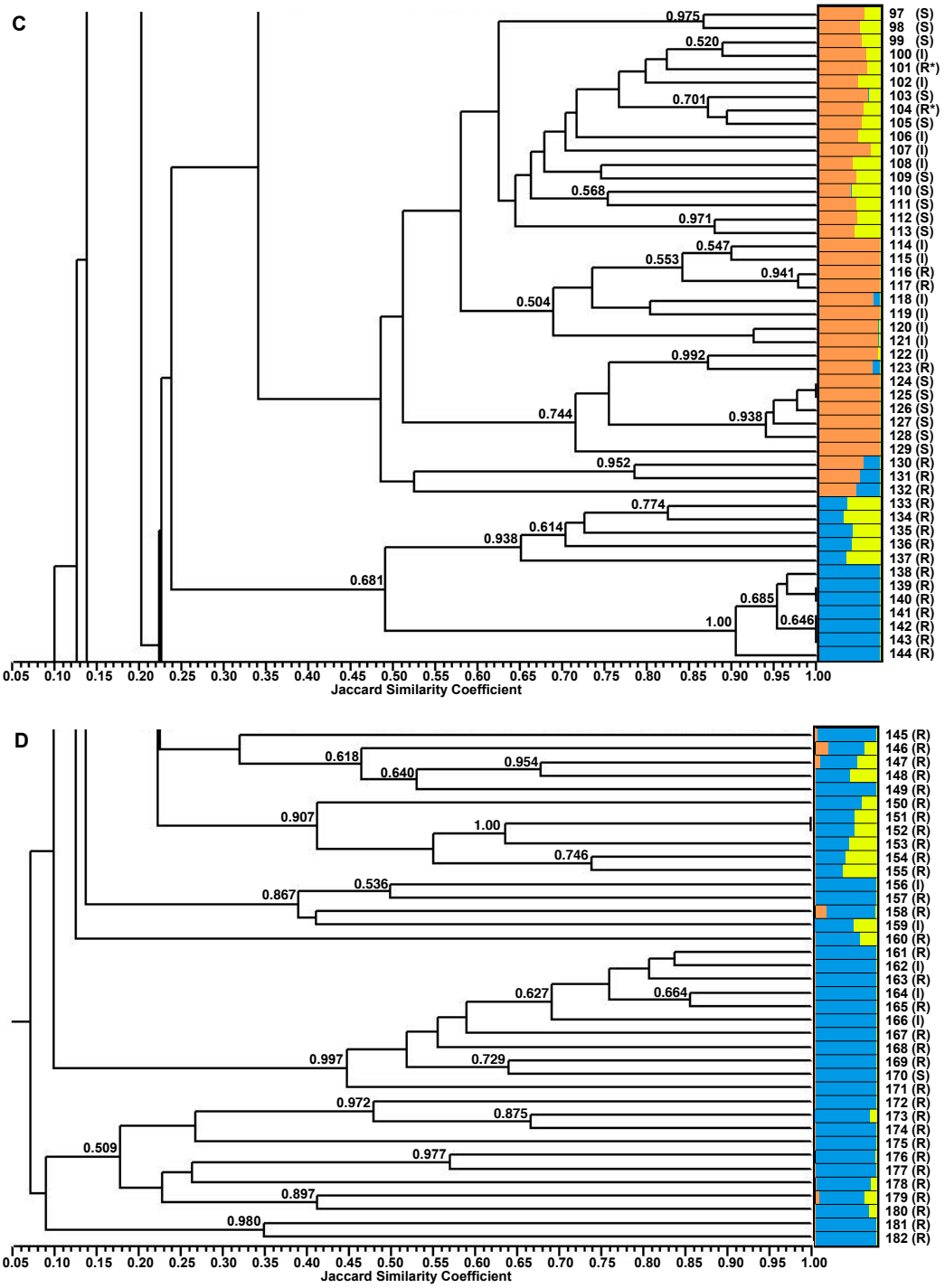


Fig 3. Unweighted pair group method using arithmetic average (UPGMA) dendrogram (left) aligned with primary model-based clustering of *Ocimum* spp. 180-accession panel and two outgroup accessions (1 and 2) using 20 EST-SSR markers. Genetic distance was calculated using the Jaccard Similarity Coefficient (x-axis) and bootstrap values are the result of a 1,000 permutations with support value greater than 0.500 shown. Three clusters (K) were inferred using Structure ver 2.2.3 and membership to each cluster is represented by proportion of yellow, orange or blue colors within

horizontal lines corresponding to each accession. Numbers right of membership histograms correspond to accessions. Response to downy mildew is indicated resistant (R) = DS < 1.0, intermediate (I) = 1.0 < DS < 3.0 or susceptible (S) = DS > 3.0. The figure is divided into four segments A (top), B, C and D (bottom).

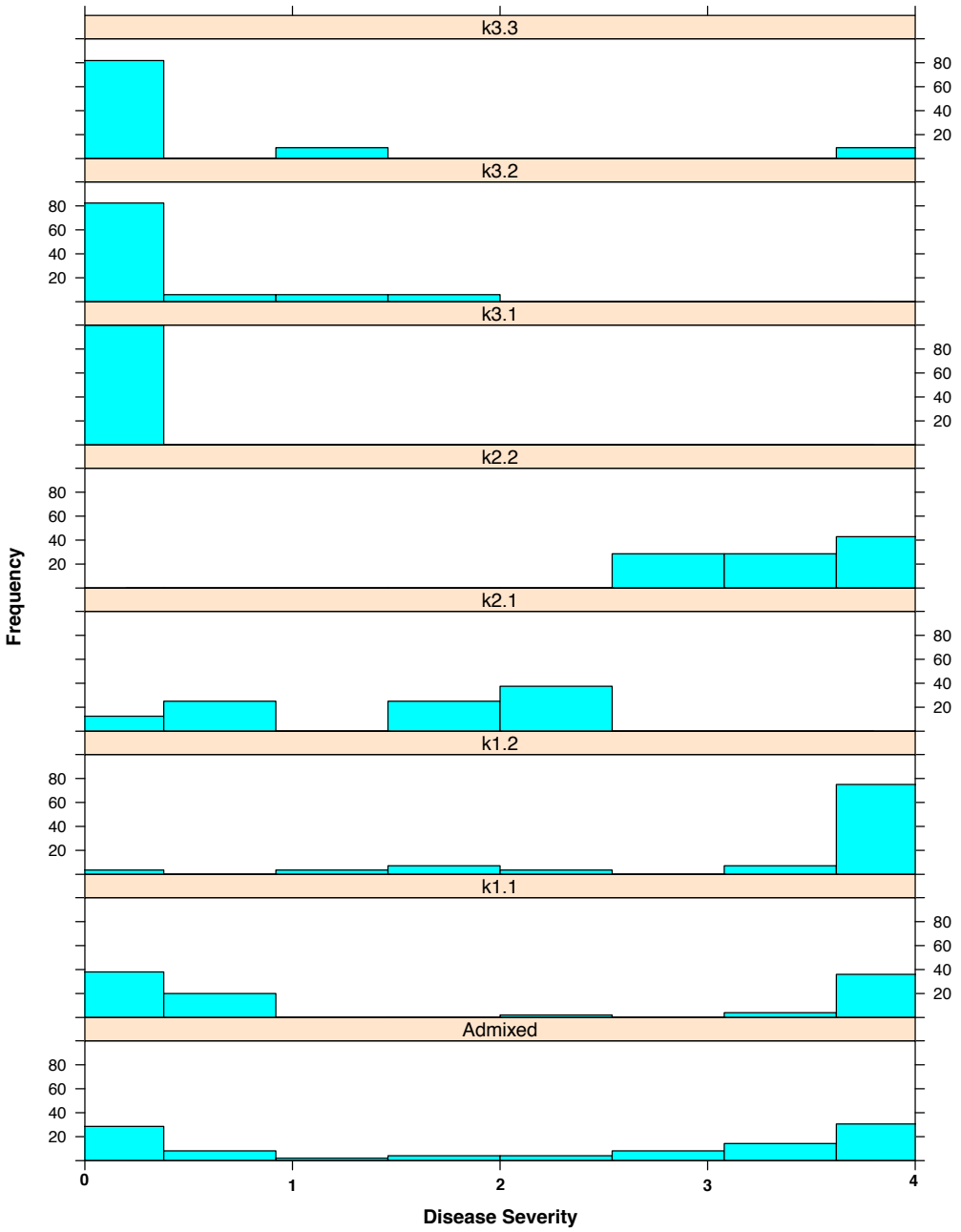
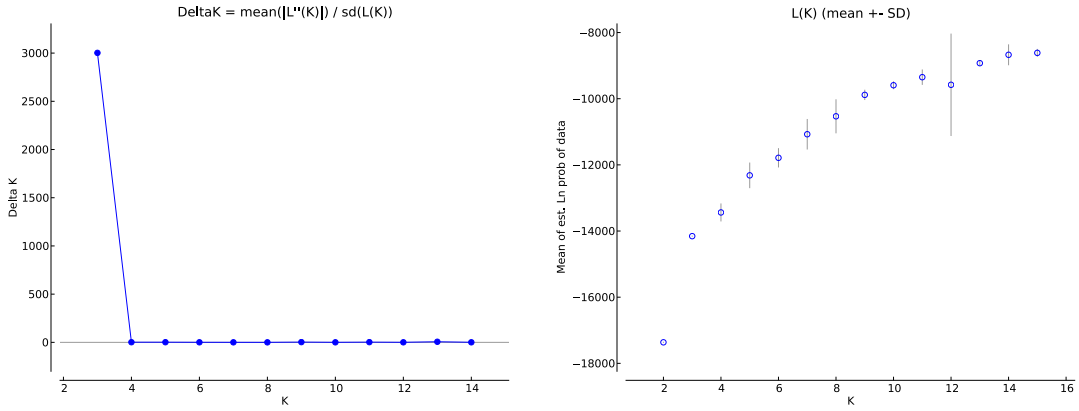


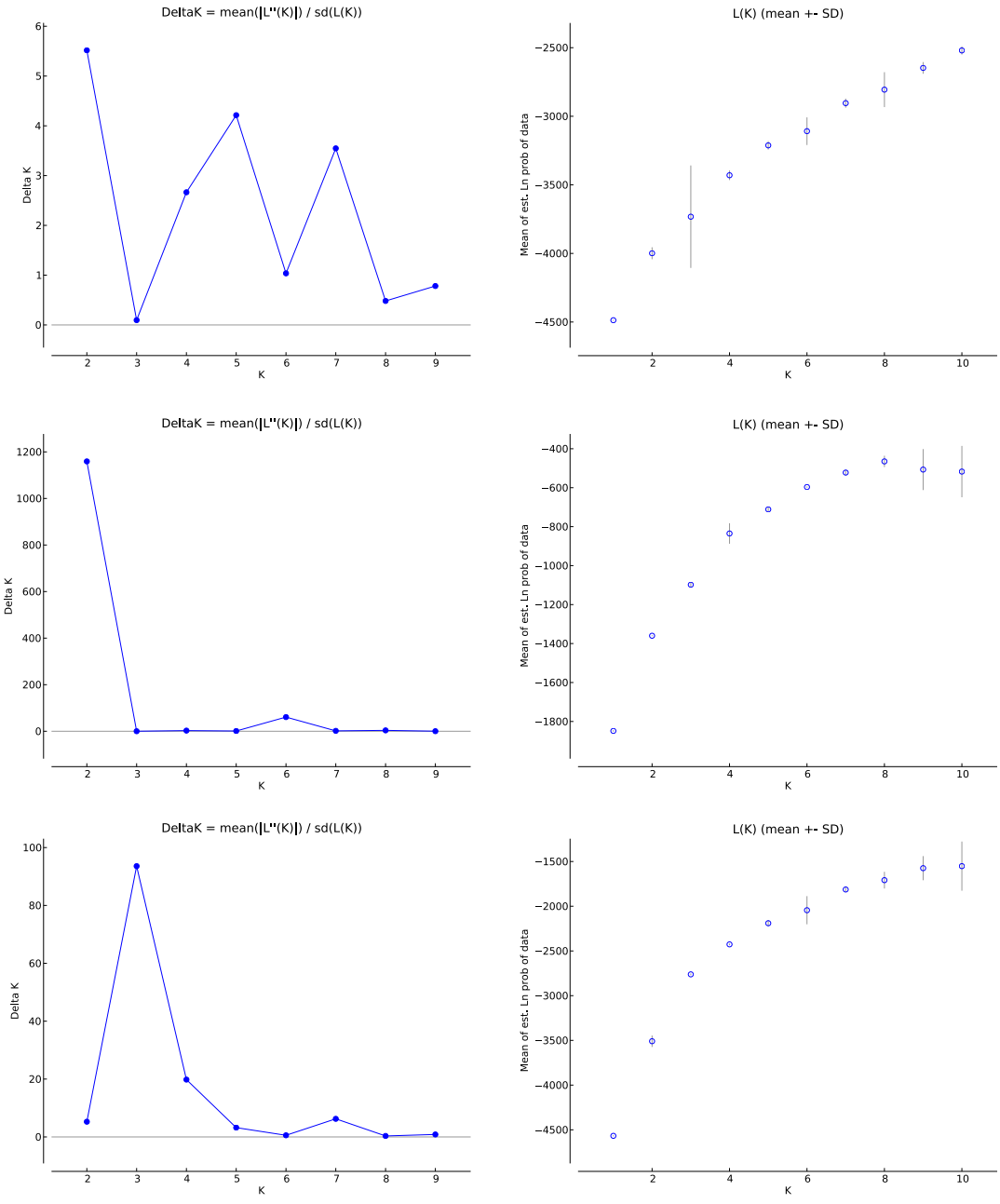
Fig. 4. Distribution of disease severity among sub-clusters derived from secondary (nested) model-based clustering analysis using Structure ver 2.2.3 software.

Supplementary Table 1. Primer sequences and melting temperatures (T_m) for the EST-SSRs used in this study.

Marker ID	NCBI / Contig ID	Forward Primer Sequence	T_m	Forward Reverse Sequence	T_m
OBNJR2sg33	DY333933	GCCTCTCCCTCCTCCATAAC	60.037	AGGCGACGAGCATGAAGTAG	60.559
OBNJR2cn29	Contig1138	TGGACATCAAATTGGCTTCA	60.049	TGGAAGGACTCGTCATCTCTC	59.395
OBNJR2sg04	DY343638	ACGATATGAGACATGGGCCT	59.39	CGCAGGTACAAGCTTCTCAA	59.22
OBNJR2sg30	DY336727	GCCAAATAATTCCTATCCGGT	59.202	CTTGGCTTTGGGAGATTCAC	59.67
OBNJR3sg124	DY331703	TAAGACAACAATCGGTGCCA	60.111	GTCCCATTCCTCCTCCGTAT	60.154
OBNJR3sg19	DY343509	AAGCCGCCCTATAAACCAAT	59.829	GGCCGTTACAAAGAGCTGAG	60.015
OBNJR3cn298	Contig2510	GACGCACCTCAAGAGTGATG	59.42	ATGGAACCATGGGAAGATGA	60.135
OBNJR3cn359	Contig2911	TAGGTCAGCTAGTGCGCAGG	61.654	GCAGAAGCGTATACATGCGA	60.006
OBNJR3cn362	Contig2969	GAAGAGATGGCTGGTCTTGG	59.803	AGACAGAGAGAGGGCAGCAG	59.883
OBNJR3sg155	DY325572	AATTCTCAGCAGGGTTGGTG	60.111	CAGCCTATTCGACGACAACA	59.864
OBNJR3sg168	DY323726	TTTATGAGATTGGCGCACAC	59.694	GCCATGTCCAGATCCTTGTT	59.934
OBNJR3sg113	DY335879	CAGATGACCACACCGAAATG	59.96	TGCAGTGAGAATGAAGGTGC	59.992
OBNJR3cn56	Contig582	GAAACAACATCCCTCATGCC	60.326	TTGAGATTGGGTTGGAGGAG	60.042
OBNJR3cn74	Contig715	GAAGGCGCTGAGAAGAAGAA	59.836	CCAATTCAACACAACCATCG	59.816
OBNJR3sg145	DY328393	GGAAATGTGGTCGTTATTCACA	59.73	CCAAAGGAAGCGACAATGAT	60.074
OBNJR3cn03	Contig100	CCGCTCTGATCTTCACTTCC	59.95	TTCACAGTCGATTCAGCAGG	59.984
OBNJR3cn210	Contig1890	CGATCATGGTGTCTCAGTGTG	60.17	GAGAGGTATCCGGTGCACAT	59.957
OBNJR3cn240	Contig2142	AACACACAAAGATCCAAACCC	58.803	TTATTTCCCAACCCACTACCA	59.188
OBNJR3sg13	DY344184	CAACAACAACGGAGAGCAAA	59.881	AAATACTTGCCGTTTGCGAC	60.138
OBNJR4cn17	Contig2461	TGAAGGCTTTGAAGAGGTAAAGA	59.548	TTCTGCTGGGCTTTGAGTTT	59.993



Supplementary Fig. 1. Structure Harvester output for primary model-based clustering using Structure ver 2.3.4 software. Plot of Δ*K* (left) and LnP(D) (right) for *K* = 1 – 15.



Supplementary Fig. 2. Structure Harvester output for secondary (nested) model-based clustering using Structure ver 2.3.4 software. Plots of Δ*K* (left) and LnP(D) (right) for *K* = 1 – 10. Three nested analyses shown for k1 (top), k2 (middle) and k3 (bottom) primary clusters.

CHAPTER 3.

Inheritance of resistance to downy mildew in sweet basil

ABSTRACT

Sweet basil (*Ocimum basilicum*) is one of the most economically important culinary herbs in the world, yet global production has become increasingly challenging due to the destructive disease downy mildew (*Peronospora belbahrii*). Although multiple sources of resistance have been identified, there are no resistant sweet basil cultivars with a commercially acceptable chemotype and phenotype available. The commercial basil cultivar Mrihani (MRI) was identified as resistant and crossed with a Rutgers susceptible sweet basil inbred line (SB22) to generate a full-sibling family. To determine the mode of inheritance for resistance to downy mildew in basil, six related generations of the MRI x SB22 family were evaluated using a Disease severity index (DSI) at northern and southern New Jersey locations over 2 years. Segregation ratios in the F_2 and backcross to the susceptible parent (BCP1) generations demonstrated chi-square goodness of fit to the two gene complementary (F_2 P =0.11; BC_1P1 P=0.04) and recessive epistatic (F_2 P=0.03; BC_1P1 P=0.63) models. Further analyses of gene effects using a weighted six-parameter scaling test provided evidence that nonallelic additive x additive and additive x dominant gene effects were highly significant (P<0.001) and resistance-reducing. All siblings in the F_1 and BC_1P2 generations were resistant (0.33>DSI) providing strong evidence that inheritance of resistance from MRI was conferred by dominant alleles. This is the first report of heritable genetic resistance that can be introduced to sweet basil without the issue of sterility barriers. Plant breeding strategies using the MRI x SB22 family should

exploit dominant gene action and remove recessive, resistance-reducing alleles from the population.

Introduction

Basil (*Ocimum* sp.) is a highly diverse genus [1] that includes at least 64 species [2], several of which are among the most widely cultivated and consumed herbs worldwide [3]. In the United States, Israel and Europe, commercial sweet basil is readily identified by a distinct phenotype and chemotype used in the fresh, dried, and culinary industries [3,4]. Downy mildew has become a devastating disease of commercial sweet basil around the world and has severely impacted major growing regions in the U.S. since 2007 [5,6]. A genus-specific host range limits *Peronospora belbahrii* infection to basil species and excludes other members of the Lamiaceae family [7,8]. Pathogenesis occurs during extended periods of leaf wetness and high humidity, which facilitate sporangia germination and entry through host stomata [9,10]. Persistence of these environmental conditions results in profuse sporulation from the abaxial leaf surface and rapid disease progress leading to complete crop loss [5]. The capacity of *P. belbahrii* to persist on both seed [11] and transplants [12] provides versatility to this species of downy mildew, which has facilitated its worldwide dissemination [9,13–16].

Conventional and organic control methods have been explored and several conventional fungicides have resulted in reduced disease severity; however, organic controls tested to date remain ineffective [17,18]. The efficacy of chemical control agents is dependent upon many factors including host developmental stage, application rate/interval, environmental conditions and inoculum levels. The foliar application of pesticides in sweet basil has a number of disadvantages including the chemical residuals [19], pathogen selection pressure [20], and increased resources required for repeated applications. The existence of *P. belbahrii* races has not been reported. However, recent

discovery of oospores [20] suggests the potential for sexual reproduction and eventual evolution of new pathogen races. Integration of downy mildew resistant sweet basil cultivars into pest management strategies represents a more sustainable control strategy that is advantageous to environmental, public health, and economic interests.

Host resistance to basil downy mildew was first reported in 2010 from field evaluations in which all *O. basilicum* cultivars and breeding lines were highly susceptible [6]. Sporulation was reduced in multiple cultivars of *O. xcitriodorum*, while sporulation and chlorosis were absent on leaves of three similar cultivars of *O. americanum* [6]. To confirm these field observations and identify new sources of resistance and tolerance to basil downy mildew, a rapid screening protocol using a controlled environment was developed [21].

Greenhouse evaluations of the United States Department of Agriculture- National Plant Germplasm System (USDA-NPGS) germplasm demonstrated extensive susceptibility among *O. basilicum* accessions with the exception of accessions PI 172996, PI 172997, and PI 172998. Sporulation was not observed on leaves of these accessions. However, presence of disease symptoms (i.e., chlorosis and necrosis) in greenhouse screenings [21] and presence of sporulation in field evaluations (J.E. Simon, unpublished) diminished the breeding value of these lines. Greenhouse screening has been essential to the identification of multiple resistant *Ocimum* sp. accessions such as PI 652053 [21,22]. However, these accessions are different species from *O. basilicum* and breeding is limited by sexual incompatibility and F1 hybrid sterility. The commercial cultivar Mrihani (Horizon Seed Co., Williams, OR) (*Ocimum* sp.) is a methyl chavicol chemotype with highly serrate and undulate leaves that was initially identified as downy mildew resistant

using the rapid screening approach [21]. Despite a substantial difference in aroma, flavor, and phenotype, this cultivar is sexually compatible with sweet basil and could be used to facilitate the development of subsequent generations from a fertile F_1.

For disease resistance breeding strategies to be most effective they must be supported by inheritance studies that measure disease response across multiple generations, environments, and years [23]. Such studies can be used to determine number of genes involved and elucidate gene action, which is essential to the adoption of an effective breeding method. True leaf downy mildew resistance in *Brassica* species has been attributed to a single dominant gene in multiple reports [24–28]. Carlsson et al. [29] determined a single recessive gene to be responsible for resistance in cotyledons of *B. oleraceae*. Traits controlled by Mendelian genes are advantageous to plant breeders because rapid gain can be achieved with fewer selection cycles. However, in many cases gene action is complicated by multiple genes acting independently or interacting in a nonallelic or epistatic fashion [30–32].

Peronospora belbahrii is an emerging pathogen of *Ocimum* sp. and inheritance of resistance has not been characterized. When prior knowledge of gene action is unavailable it is important that an appropriate mating design is selected to capture as much genetic variation as possible through use of multiple populations from the same or different parents [32]. The development of six related generations derived from two inbred lines with differential response to a particular trait of interest provides a metrical system for partitioning of gene effects using scaling tests [30,32]. This classic mating design has proven effective in determinations of heritability and gene action when characterizing new sources of disease resistance [33,34]. The objective of this study was

to investigate the mode of inheritance for resistance to downy mildew in a segregating full-sibling sweet basil family. Results will provide the basis by which to design the appropriate breeding and selection strategies for the development of downy mildew resistance in sweet basil.

Materials and Methods

Plant Material & Controlled Crosses

The two parent genotypes in this study were selected based on a consistent, differential response to downy mildew through repeated field and greenhouse evaluations at Rutgers University greenhouses. Inbred sweet basil breeding line SB22 (Fig. 1A) was found to be highly susceptible, while the cultivar Mrihani (Fig. 1B) (Horizon Seed Co., Williams, OR) demonstrated resistance with no sporulation and little or no chlorosis. The latter genotype was therefore selected as the resistant parent and designated MRI following two generations of self-pollination. Lacking any prior knowledge with regard to the genetic basis by which MRI confers resistance, a six-generation mating design consisting of susceptible parent (SB22) (Fig 1A), resistant parent (MRI) (Fig 1B), F_1 (maternal parent = MRI) (Fig. 1C), F_1R (maternal parent = SB22), F_2 (Fig. 1H-K), backcross to SB22 (BC_1P1) (Fig. 1D-G) and backcross to MRI (BC_1P2) generations were selected to allow for measurement of potential additive, dominant and epistatic gene effects [32]. Crosses were made in early morning hours by emasculation of unopened (immature) flower whorls followed by hand-pollination 24 and 48 h later. Glassine bags (Seedburo, Des Plaines, IL) were used to cover emasculated flowers in order to prevent any incidental outcrossing from taking place. In total, 30 susceptible parent (SB22), 30 resistant parent (MRI), 30 F_1, 30 F_1R, 300 F_2, 180 BC_1P1 and 144 BC_1P2 individuals

were used in this study. Seeds were germinated on blotting paper moistened with sterile distilled water and incubated at 30/20 ºC corresponding to a 12 h light/dark schedule. Seedlings were transferred to a 72- or 128-celled flat with growing mix (Fafard Grow Mix 2; Sun Gro Horticulture, Agawam, MA) upon emergence of cotyledons and maintained under greenhouse conditions at the Rutgers Research Greenhouses in New Brunswick, NJ until field establishment.

Field Establishment and Experimental Design

This study was conducted at two locations over 2 years. In year 1 an exploratory field experiment was initiated in which five generations: SB22, MRI, F_1, F_1R, F_2 and BC_1P1 were evaluated at a single site, the Rutgers Agricultural Research and Extension Center (RAREC) in Bridgeton, NJ (southern New Jersey). In year 2, this experiment was repeated with six generations: SB22, MRI, F_1, F_1R, F_2, BC_1P1 and BC_1P2 from new crosses at RAREC and at a second location, the Snyder Research and Extension Farm in Pittstown, NJ (northern New Jersey). The two locations in this study are separated by approximately 118 km and are known to have different environmental conditions (i.e., temperature and rainfall patterns). RAREC is located in the outer coastal plains of southern New Jersey in a sandy-loam soil type, while the northern location has a silt-loam soil type. All genotypes evaluated in year two were cloned by vegetative cuttings, with one clone per genotype represented at each of the two locations to provide a measure of genotype by environment interaction. Clones were transplanted in single rows with 1-m spacing at both locations to allow for single plant evaluations.

All field experiments were conducted in raised beds with 0.032- mm plastic mulch spaced on 1.5- m centers and with drip irrigation applied as needed. Fertilization,

insecticides, and herbicides were applied as needed and according to local standard production practices at each location in both years of the study. Field experiments were arranged in a randomized complete block design with progeny of each generation was divided equally into three blocks. Thus, each block contained 10 plants per parent, 10 plants per F_1, 10 plants per F_1R, 100 plants per F_2, 60 plants per BC_1P1 and 48 plants per BC_1P2.

Inoculation and Disease Rating

Peronospora belbahrii inoculum was consistently present at the northern and southern New Jersey locations in both years of this study and was consistent with other related studies on basil downy mildew conducted at these sites [6,17]. Overhead irrigation was regularly applied to maintain leaf wetness and relative humidity during extended dry periods. Control plots and guard rows consisting of highly susceptible *O. basilicum* commercial cultivar DiGenova were included within each block to ensure uniform inoculum and aid in the determination of appropriate rating dates.

To ensure that the response variable was strictly a measurement of reaction to *P. belbahrii*, evaluations were made on the basis of percent sporulation. A five-point ordered categorical scale (0-4) was used in which 0=no sporulation, 1=1% to 10% sporulation, 2=11% to 25%, 3=26% to 50% and 4=51% to 100%. This scale facilitated rapid scoring of multiple leaves from individual plants, while providing a repeatable and representative measure of disease reaction on an individual plant basis. Six mature leaves were detached from each plant and assigned a score from which a disease severity index (DSI) [19,28,35] was calculated on a single-plant basis using the equation:

$$DSI = \frac{\sum(\text{single leaf x disease rating})}{(\text{number of leaves scored x maximum disease rating})}$$

Disease progress was monitored using susceptible check plots. Initial disease evaluations were made upon observance of sporulation on all check plots and subsequent evaluations were made with increasing disease severity. Disease ratings were made on three dates (8 Aug., 26 Aug., and 15 Sept.) in 2013 and four dates (25 Aug. and 10 Sept. in southern New Jersey; 28 Aug. and 16 Sept. in northern New Jersey) in 2014 for a total of seven disease rating over two years. For each year x location combination, disease ratings data corresponding to the highest family mean DSI, or greatest disease selection pressure, was selected for subsequent statistical analysis. These ratings were performed 33 d after planting (DAP) in southern New Jersey in 2013, 49 DAP in southern New Jersey in 2014, and 53 DAP in northern New Jersey in 2014.

Statistical Analysis

Means, variances, and frequency distributions of DSI scores among generations were calculated separately for each of the three rating dates selected over two years using PROC UNIVARIATE (SAS version 9.4; SAS Institute, Cary, NC). Bartlett's homogeneity of variances tests were performed using the DSI scores generation to determine whether it was appropriate for data to be pooled by year, location, and blocks [36,37]. The non-segregating F_1 and F_1R generation means were subjected to a two-sample t test of independence for detection of maternal effects on disease resistance [37].

An analysis of variance (ANOVA) was performed using PROC MIXED (SAS version 9.4) in which location and block effects were random, but generation was fixed because purposeful crosses were performed using parents with a previously observed disease response. Unbiased estimates of variance were calculated for the effect of location (σ^2_l), location x generation (σ^2_{lg}) and location x block (σ^2_{lb}) using the restricted

maximum likelihood method. A mean separation was performed to determine differences among generations using the least significant difference ($LSD_{0.05}$) method.

To provide an estimation for the number of loci involved in disease response, five separate formulas were used to determine the minimum number of effective factors using the SASQuant 1.3 program (SAS, version 9.4) [38]. Five separate methods including Wright's method [37], Mather's method [32] and [39] Lande's methods I-III provided five estimates from which an overall mean was calculated.

Chi-square tests were used to determine goodness of fit to hypothesized gene models based on observed and expected number of resistant and susceptible individuals in the F_2 and BC_1P1 generations. Based on comparisons of DSI scores with observed plant disease reactions in the field and analysis of generation frequency distributions, plants with a DSI score < 0.33 (slightly more than 10% sporulation) were considered resistant. A portion of the population demonstrated a disease response within a DSI interval 0.33 to 0.66, which suggested an intermediate resistance reaction among these plants. In a similar study, Tetteh et al. [34] identified multiple hypothesized segregation ratios that could be fit to observed segregation ratios with and without an intermediate response. This approach was also used in this study by performing chi-square tests to determine goodness of fit to multiple hypothesized major gene models with the observed number of resistant, intermediate and susceptible plants. In gene models that included an intermediate disease response (one incomplete dominant gene; two additive genes; two genes with recessive epistasis), a DSI > 0.66 (more than 40% sporulation) was considered downy mildew susceptible. In models that did not include an intermediate response (one

dominant gene, two complementary genes, two genes with dominant suppression epistasis), a DSI > 0.33 was considered susceptible.

Addition of the BC_1P2 generation in the 2014 field experiment permitted estimation of the main and epistatic gene effects affecting response to downy mildew. Gene effects were partitioned into six related parameters using a weighted least square regression of the extended joint scaling test [30]. Genetic parameters included the mean (m), additive (d), dominant (h), additive x additive (i), additive x dominant (j) and dominant x dominant (l) effects. Standard errors and student's t significance level were calculated for all gene effects using the SAS Quant 1.3 program (SAS, version 9.4) [38].

Results and Discussion

The 2013 and 2014 results (Table 1) indicate the importance of multi-year evaluations when determining disease response to *P. belbahrii*, a pathogen in which virulence is highly dependent upon environmental conditions. Both the 2013 F_2 variance and mean were significantly lower when compared to the F_2 data at either 2014 location (Table 1). A test of homogeneity of variances demonstrated a highly significant difference among F_2 data across years (χ^2= 15.28; P<0.001). Graphical analysis of the 2013 DSI frequency distribution and generation means revealed greater skewness towards resistance (data not shown) in comparison with the 2014 DSI distributions (Fig. 2) providing evidence that downy mildew pressure was significantly higher in 2014 than 2013. Estimates of heritability can be skewed and segregation ratios distorted under conditions of insufficient selection pressure [32]. In the absence of adequate disease pressure, susceptible genotypes could be mistakenly selected as resistant and substantially reduce the effectiveness of plant breeding strategies. To provide the most

representative measure of disease response, 2014 data was selected for subsequent statistical analysis of generation means, variances and distributions. The F_2 variances in 2014 were not significantly different when compared across blocks (χ^2= 2.92; *P*=0.23) providing evidence that variation in disease response (phenotypic variance) could be achieved using 100 F_2 individuals. An additional test for homogeneity of F_2 variances among locations demonstrated no significant difference (χ^2= 0.50; *P*=0.48) permitting 2014 data from northern New Jersey and southern New Jersey to be pooled in subsequent statistical analyses.

Generation means and variances. Significant differences (*P*<0.001) were observed among generations of the MRI x SB22 family (Table 2). Mean DSI among parents demonstrated a differential response to downy mildew (SB22 DSI =0.967; MRI DSI =0.046). This is the first report of resistance in genotype MRI, which was confirmed in this study in 2 years and two locations (Table 1). Although multiple downy mildew resistant basil accessions have been reported [6,21,22] these genotypes are genetically divergent or taxonomically distinct from *O. basilicum*. Hybridization among *O. basilicum* species can be achieved through purposeful cross-pollination [40]. Yet attempts to introgress disease resistance genes into traditional sweet basil by wide crosses or interspecific hybridization have been obstructed by sterility barriers (data not shown). These issues are likely due to significant differences in genome size and basic chromosome numbers in susceptible *O. basilicum* accessions and resistant basil species including *O. americanum*, *O. kilimandscharicum*, *O. gratissimum* and *O. tenuiflorum* [41,42]. Identification of resistance in MRI provides a downy mildew resistant parent that

is fully cross-compatible with commercial sweet basil, suggesting similar genome constitution.

The F_1 and F_1R generations were resistant (DSI<0.33) and not significantly different (P<0.335) according to a two-sample t test (Table 1), indicating that the resistance conferred by MRI is not maternally inherited and reciprocal F_1 generations could be pooled. A strong dominant effect was demonstrated by an F_1 mean (0.137) substantially lower than the mid-parent value of 0.506 and not significantly different from the MRI or the BC_1P2 mean (Table 3). The absence of segregation in the BC_1P1 generation suggested that MRI confers resistance through a dominant gene action. Both the F_2 and BC_1P1 generations demonstrated segregation for disease response with distributions that spanned from the lowest (0.0) to highest (1.0) possible DSI scores (Fig. 2). The BC_1P1 DSI mean was the highest among segregating generations and significantly greater (P<0.05) than the next highest F2 generation mean (Table 3). The F_2 DSI frequency distribution exhibited skewness towards resistance, while the BC_1P1 distribution showed a substantial shift towards susceptibility (Fig. 2). Within the BC_1P1 generation 44 genotypes (plants) or 24.4% were considered resistant (DSI<0.33), while 155 genotypes or 51.6% of the F_2 generation were resistant.

Heritability

ANOVA was used to calculate all variation due to environment. ANOVA provided evidence that block, block by location, and generation by location effects were not significant (Table 2). Although the effect of location was significant at a P<0.05 confidence level, the variance component estimate of 0.001 was quite low relative to the segregating F_2 and BC_1P1 generations. Heritability can be described simply as the

proportion of genetic variance (broad sense) or additive genetic variance (narrow sense) responsible for measured phenotypic variance, which includes all genetic and environmental variances [23]. In this study, the effect of location was a significant variance component of environment; however, it accounted for just 0.2% of the BC_1P1 and F_2 variation. These estimates provide convincing evidence that the observed phenotypic variance across locations were largely attributed to genotypic variance. This is advantageous to breeding strategies, indicating that downy mildew resistance in the MRI x SB22 family is heritable across multiple locations given adequate disease pressure.

Gene action

The presence of distinctly resistant, intermediate and susceptible responses, suggests that downy mildew resistance in the MRI x SB22 family was under major gene control. Observed segregation ratios in the F_2 and BC_1P1 populations were first tested for chi-square goodness-of-fit for monohybrid gene models (Table 4). Both single dominant and incomplete dominant expected ratios models were rejected at the $P<0.01$ confidence level, suggesting more than one locus was involved in the response to downy mildew. The number of effective factors, or loci, involved in response was estimated to be 1.4 (Wright's method), 5.2 (Mather's method), 1.1 (Lande's method I), 1.3 (Lande's Method II) and 0.9 (Lande's Method III). These estimates yielded an average of 2.0 and supported the major gene control hypothesis as well as the lack of evidence for single gene control. While imprecise, these five estimates are useful in detecting control of resistance by a few major genes [43]. Multiple dihybrid gene models were evaluated and a chi-square test failed to reject the hypothesis for the two gene complementary (F_2

P=0.11); BC_1P1 *P*=0.04) and recessive epistasis models (F_2 P=0.03; BC_1P1 P=0.63) (Table 4). Both gene models suggest that digenic epistasis is involved in resistance to basil downy mildew and a simple additive-dominant model is not adequate for describing the gene action. These results demonstrated that an extended scaling test was necessary for estimating significant nonallelic effects, which contributed to the variation observed in the MRI x SB22 family.

Additive (d) and dominant (h) gene effects were both highly significant (*P*<0.001) and opposite (Table 5). A dominant, negative effect of -0.58 indicated involvement of a heterozygous resistant locus, whereas the positive additive effect of 0.56 indicated a homozygous, resistance-reducing locus. A highly significant (*P*<0.001) additive x additive (i) gene effect of 0.88 provided the strong evidence for epistasis or the interaction of these loci. A positive value for the additive x additive effect suggests that interaction between homozygous loci results in a substantial resistance-reducing effect. Similarly, the additive x dominant (j) interaction was positive and highly significant (*P*<0.001), suggesting that homozygous and heterozygous loci were also interacting in a resistance-reducing capacity. Dominant x dominant (l) effects were not significant due to comparable F_2 and BC_1P1 generation means and a relatively high standard error of 0.22 (Table 5).

Results of the joint scaling test provide support for Chi-square test hypotheses in which homozygous recessive alleles at either locus reduce the resistance response, while single dominant alleles confer resistance at a given locus. A significant and positive additive x dominant gene effect supports the recessive epistasis model in which a homozygous recessive locus masks expression of dominant resistance at a second locus.

A very high additive x additive gene effect indicates that a significant interaction exists between homozygous recessive loci, which would result in complete susceptibility within a two gene recessive epistatic model. Chi-square goodness of fit to the 9:7 ratio failed to reject the two gene complementary hypothesis, which was further supported by the proximity of the F_2 mean DSI (0.409) to the midparent mean DSI (0.506) rather than the F_1 mean (0.137). This depression in the F_2 population mean response suggests complementary epistasis [30]. Wang et al. [44] demonstrated that true leaf downy mildew resistance in a broccoli doubled-haploid line was controlled by complementary epistasis based on goodness-of-fit to expected F_2 and BC_1P1 segregation ratios.

The development of additional populations and quantitative trait loci (QTLs) is useful in providing greater resolution with regard to the interaction of loci conferring downy mildew resistance [45] and is needed to further elucidate gene action in the MRI x SB22 sweet basil family. A recent study by den Boer et al. [46] found that quantitative trait loci (QTLs) conferring field resistance to downy mildew in *Lactuca saligna* backcross inbred lines (BILs) resulted in "more-than-additive" and "less-than-additive" epistasis when intercrossed.

Epistasis has proven to be ubiquitous in plant resistance responses [47–49] and in multiple instances of resistance to downy mildew [44,45,35]. Detection of nonallelic gene effects has important implications when designing effective breeding programs. Both complementary and recessive epistatic models introduce similar complexity to breeding efforts, but can be overcome utilizing appropriate breeding and selection strategies.

Typically, dominant or "non-fixable" gene effects are exploited by hybrid breeding programs when possible [50]. In sweet basil, hybrid seed production is not a

feasible breeding approach currently due to the meticulous nature of cross-pollinations and lack of an available doubled haploid protocol. A backcross breeding method using SB22 as the recurrent parent would be advantageous in decreasing undesirable phenotypes attributed to MRI (Fig. 1B). However, a 47% decrease in disease resistant individuals was observed in the BC_1P1 generation relative to the F_2 corresponding to a 50% decrease in the population gene frequency contributed by MRI. Therefore, large backcross populations would be required to overcome the frequency of recessive resistance-reducing alleles and achieve reasonable numbers of resistant individuals. Given the gene action described in this study, pedigree breeding would be a most effective approach in the fixation of dominant, downy mildew resistant loci and the elimination of recessive alleles conferring susceptibility from the breeding population [50]. Simultaneous selection for traits associated with SB22 (Fig.1A) will yield downy mildew resistant inbred breeding lines with a commercial sweet basil phenotype and chemotype.

This is the first characterization of genetic resistance to downy mildew in basil, as well as the use of six related generations in a single family to evaluate gene action. Results demonstrate the use of this six-generation mating design was not only appropriate, but necessary for detecting the complex inheritance pattern controlling downy mildew resistance in the MRI x SB22 full sib basil family.

Conclusion

Results of this study indicate that resistance to basil downy mildew is controlled by digenic epistasis in the MRI x SB22 family. Generation means, Chi-square goodness-of-fit models and the joint scaling test support the conclusion that dominant alleles confer

resistance, while homozygous recessive alleles confer susceptibility. Breeding strategies should exploit the dominant gene action by fixing it at both loci and removing susceptibility alleles from the breeding population. This study provides evidence for a heritable form of downy mildew resistance that can be used to breed for genetic resistance in a commercial sweet basil.

References

1. Vieira RF, Goldsbrough P, Simon JE. Genetic diversity of basil (*Ocimum* spp.) based on RAPD markers. J Am Soc Hort Sci. 2003;128: 94–99.
2. Paton A, Harley MR, Harley MM. *Ocimum*: An overview of classfication and relationships. In: Hiltunen R, Holm Y, editors. Basil: The Genus *Ocimum*. Amersterdam: Harwood Academic Publishers; 1999. pp. 1–38.
3. Simon JE, Quinn J, Murray J. Basil: A Source of Essential Oils. In: Janick J, Simon JE, editors. Advances in New Crops. Portland, OR: Timber Press; 1990. pp. 484–489.
4. De Masi L, Siviero P, Esposito C, Castaldo D, Siano F, Laratta B. Assessment of agronomic, chemical and genetic variability in common basil (*Ocimum basilicum* L.). Eur Food Res Technol. 2006;223: 273–281. doi:10.1007/s00217-005-0201-0
5. Roberts PD, Raid RN, Harmon PF, Jordan SA, Palmateer AJ. First Report of Downy Mildew Caused by a *Peronospora* sp. on Basil in Florida and the United States. Plant Dis. 2009;93: 199. doi:10.1094/PDIS-93-2-0199B
6. Wyenandt CA, Simon JE, McGrath MT, Ward DL. Susceptibility of basil cultivars and breeding lines to downy mildew (*Peronospora belbahrii*). HortScience. 2010;45: 1416–1419.
7. Belbahri L, Calmin G, Pawlowski J, Lefort F. Phylogenetic analysis and real time PCR detection of a presumbably undescribed *Peronospora* species on sweet basil and sage. Mycol Res. 2005;109: 1276–1287. doi:10.1017/S0953756205003928
8. Thines M, Telle S, Ploch S, Runge F. Identity of the downy mildew pathogens of basil, coleus, and sage with implications for quarantine measures. Mycol Res. 2009;113: 532–540. doi:10.1016/j.mycres.2008.12.005
9. Garibaldi A, Bertetti D, Gullino ML. Effect of leaf wetness duration and temperature on infection of downy mildew (*Peronospora* sp.) of basil. J Plant Dis Prot. 2007;114: 6–8.
10. Koroch AR, Villani TS, Pyne RM, Simon JE. Rapid Staining Method to Detect and Identify Downy Mildew (*Peronospora belbahrii*) in Basil. Appl Plant Sci. 2013;1: 1300032. doi:10.3732/apps.1300032
11. Garibaldi a, Minuto G, Bertetti D, Gullino ML. Seed transmission of *Peronospora* sp.of Basil. J Plant Dis Prot. 2004; 465–469.
12. Djalali Farahani-Kofoet R, Römer P, Grosch R. Systemic spread of downy mildew in basil plants and detection of the pathogen in seed and plant samples. Mycol Prog. 2012;11: 961–966. doi:10.1007/s11557-012-0816-z
13. Blomquist CL, Rooney-Latham S, Nolan PA. First Report of Downy Mildew on Field-Grown Sweet Basil Caused by a *Peronospora* sp. in San Diego County, California. Plant Dis. 2009;93: 968. doi:10.1094/PDIS-93-9-0968A
14. Kanetis L, Vasiliou A, Neophytou G, Samouel S, Tsaltas D. First Report of Downy Mildew Caused by Peronospora belbahrii on Sweet Basil (*Ocimum basilicum*) in Cyprus. Plant Dis. 2013;98: 283. doi:10.1094/PDIS-07-13-0759-PDN
15. McLeod A, Coertze S, Mostert L. First Report of a *Peronospora* Species on Sweet Basil in South Africa. Plant Dis. 2006;90: 1115. doi:10.1094/PD-90-1115A
16. Nagy G, Horváth A. Occurrence of Downy Mildew Caused by *Peronospora*

belbahrii on Sweet Basil in Hungary. Plant Dis. 2011;95: 1034. doi:10.1094/PDIS-04-11-0329

17. Homa K, Ward DL, Wyenandt CA, Simon JE. Evaluation of Fungicides for the Control of *Peronospora belbahrii* on Sweet Basil in New Jersey. Plant Dis. 2014;98: 1561–1566. doi:10.1094/PDIS-02-14-0200-RE
18. Mersha Z, Zhang S, Raid RN. Evaluation of systemic acquired resistance inducers for control of downy mildew on basil. Crop Prot. Elsevier Ltd; 2012;40: 83–90. doi:10.1016/j.cropro.2012.04.031
19. Gilardi G, Demarchi S, Garibaldi A, Gullino ML. Management of downy mildew of sweet basil (*Ocimum basilicum*) caused by *Peronospora belbahrii* by means of resistance inducers, fungicides, biocontrol agents and natural products. Phytoparasitica. 2013;41: 59–72. doi:10.1007/s12600-012-0264-y
20. Cohen Y, Vaknin M, Ben-Naim Y, Rubin AE, Galperin M, Silverman D, et al. First Report of the Occurrence and Resistance to Mefenoxam of *Peronospora belbahrii*, Causal Agent of Downy Mildew of Basil (*Ocimum basilicum*) in Israel. Plant Dis. 2013;97: 692. doi:10.1094/PDIS-12-12-1126-PDN
21. Pyne RM, Koroch AR, Wyenandt CA, Simon JE. A Rapid Screening Approach to Identify Resistance to Basil Downy Mildew (*Peronospora belbahrii*). HortScience. 2014;49: 1041–1045.
22. Djalali Farahani-Kofoet R, Romer P, Grosch R. Selecting basil genotypes with resistance against downy mildew. Sci Hortic. 2014;179: 248–255. doi:10.1016/j.scienta.2014.09.036
23. Holland JB, Nyquist WE, Cervantes-Martinez CT. Estimating and Interpreting Heritability for Plant Breeding: An Update. In: Janick J, editor. Plant Breeding Reviews. John Wiley & Sons, Inc.; 2003.
24. Farnham MW, Wang M, Thomas CE. A single dominant gene for downy mildew resistance in broccoli. Euphytica. 2002;128: 405–407. doi:10.1023/A:1021279426050
25. Jensen BD, Hockenhull J, Munk L. Seedling and adult plant resistance to downy mildew (*Peronospora parasitica*) in cauliflower (*Brassica oleracea* convar. *botrytis* var. *botrytis*). Plant Pathol. 1999;48: 604–612. doi:10.1046/j.1365-3059.1999.00388.x
26. Monteiro AA, Coelho PS, Bahcevandziev K, Valério L. Inheritance of downy mildew resistance at cotyledon and adult-plant stages in “Couve Algarvia” (*Brassica oleracea* var. *tronchuda*). Euphytica. 2005;141: 85–92. doi:10.1007/s10681-005-5696-8
27. Natti JJ, Dickson MH, Atkin JD. Resistance of *Brassica oleracea* varieties to downy mildew. Phytopathology. 1967;57: 144–147.
28. Mahajan V, Gill HS, More TA. Inheritance of downy mildew resistance in Indian cauliflower (group III). Euphytica. 1995;86: 1–3. doi:10.1007/BF00035932
29. Carlsson M, von Bothmer R, Merker A. Screening and evaluation of resistance to downy mildew (*Peronospora parasitica*) and clubroot (*Plasmodiophora brassicae*) in genetic resources of *Brassica oleracea*. Hereditas. 2004;141: 293–300. doi:10.1111/j.1601-5223.2004.01818.x
30. Hayman B. Separation of epistatic from additive and dominance variation in generation means. Heredity.1958;12: 371–390.

31. Jinks JJ, Jones RM. Estimation of the components of heterosis. Heredity. 1957;43: 223–234.
32. Mather K, Jinks JL. Biometrical genetics: The study of continuous variation. 2nd ed. London, UK: Chapman and Hall; 1971.
33. Tetteh AY, Wehner TC, Davis AR. Inheritance of resistance to the new race of powdery mildew in watermelon. Crop Sci. 2013;53: 880–887. doi:10.2135/cropsci2012.07.0453
34. Tetteh AY, Wehner TC, Davis AR. Inheritance of Resistance to Powdery Mildew Race 2 in *Citrullus lanatus* var. *lanatus*. HortScience. 2013;48: 1227–1230.
35. Shashikumar KT, Pitchaimuthu M, Rawal RD. Generation mean analysis of resistance to downy mildew in adult muskmelon plants. Euphytica. 2010;173: 121–127. doi:10.1007/s10681-010-0132-0
36. Ostle B, Malone LC. Statistics in Research. Ames, IA: Iowa State University Press; 1988.
37. Wright JN. The genetics of quantitative variability. In: Wright S, editor. Evolution and genetics of populations. 2nd ed. Chicago, IL: University of Chicago Press; 1968.
38. Gusmini G, Wehner TC, Donaghy SB. SASQuant: A SAS software program to estimate genetic effects and heritabilities of quantitative traits in populations consisting of 6 related generations. J Hered. 2007;98: 345–350. doi:10.1093/jhered/esm033
39. Lande R. The minimum number of genes contributing to quantitative variation between and within populations. Genetics. 1981;53: 541–553.
40. Putievesky E, Paton A, Lewinsohn E, Ravid U, Haimovich D, Katzir I, et al. Crossability and relationship between *Ocimum basilicum* L. J Herbs, Spices, Med Plants. 1999;6: 11–24.
41. Koroch AR, Wang W, Michael TP, Dudai N, Simon JE, Belanger FC. Estimation of nuclear DNA content of cultivated *Ocimum* species by using flow cytometry. Isr J Plant Sci. 2010;58: 183–189. doi:10.1560/IJPS.59.3-4.183
42. Paton AJ, Putievesky E. Taxonomic problems and cytotaxonomic relationships between and within varieties of *Ocimum basilicum* and related species (Labiateae).. Kew Bull. 1996;51: 509–524.
43. Kozik EU, Klosinska U, Call AD, Wehner TC. Heritability and genetic variance estimates for resistance to downy mildew in cucumber accession ames 2354. Crop Sci. 2013;53: 177–182. doi:10.2135/cropsci2012.05.0297
44. Wang M, Farnham MW, Thomas CE. Inheritance of True Leaf Stage Downy Mildew Resistance in Broccoli. J Am Soc Hortic Sci. 2001;126: 727–729.
45. Perchepied L, Bardin M, Dogimont C, Pitrat M. Relationship between loci conferring downy mildew and powdery mildew resistance in melon assessed by quantitative trait Loci mapping. Phytopathology. 2005;95: 556–565. doi:10.1094/PHYTO-95-0556
46. den Boer E, Pelgrom KTB, Zhang NW, Visser RGF, Niks RE, Jeuken MJW. Effects of stacked quantitative resistances to downy mildew in lettuce do not simply add up. Theor Appl Genet. 2014;127: 1805–1816. doi:10.1007/s00122-014-2342-7
47. Bnejdi F, Saadoun M, Allagui MB, El Gazzah M. Epistasis and heritability of

resistance to Phytophthora nicotianae in pepper (*Capsicum annuum* L.). Euphytica. 2009;167: 39–44. doi:10.1007/s10681-008-9857-4

48. Fakhfakh MM, Yahyaoui A, Rezgui S, Elias EM, Daaloul A. Inheritances of fusarium head blight resistance in a cross involving local and exotic durum wheat cultivars. Crop Sci. 2011;51: 2517–2524. doi:10.2135/cropsci2011.03.0140
49. Keri M, Berg CGJ Van Den, Mcvetty PBE, Rimmer SR. Inheritance of Resistance to *Leptosphaeria maculans* in *Brassica juncea*. 1997;3043: 594–598.
50. Fehr WR. Principles of Cultivar Development. New York, NY: MacMillan Publishing Company; 1987.

Table 1. Means, variances and number of genotypes among generations of the basil MRI x SB22 full-sib family evaluated for response to downy mildew over two years at northern New Jersey and southern New Jersey locations.

		Generation mean[z]						
Year	Location	MRI[y]	SB22[x]	F_1[w]	F_1R[v]	F_2	BC_1P1[u]	BC_1P2[t]
2013	southern New Jersey	0.000	0.986	0.104	0.133	0.265	0.575	-
2014	southern New Jersey	0.042	0.958	0.086	0.096	0.381	0.607	0.057
2014	northern New Jersey	0.051	0.975	0.176	0.190	0.438	0.670	0.096
		Generation variance[s]						
Year	Location	MRI	SB22	F_1	F_1R	F_2	BC_1P_1	BC_1P_2
2013	southern New Jersey	0.000	0.001	0.001	0.003	0.070	0.088	-
2014	southern New Jersey	0.002	0.002	0.005	0.004	0.109	0.128	0.006
2014	northern New Jersey	0.001	0.001	0.002	0.003	0.100	0.110	0.011
		Genotypes (no./generation)						
Year	Location	MRI	SB22	F_1	F_1R	F_2	BC_1P_1	BC_1P_2
2013	southern New Jersey	30	30	30	30	300	181	-
2014	southern New Jersey	30	30	30	30	300	180	144
2014	northern New Jersey	30	30	30	30	300	180	144

[z]Mean and variance estimates generated from Disease severity index (DSI) scores of six mature leaves from individual plants.

[y]resistant parent (P2); [x]susceptible parent (P1); [w]maternal parent is SB22 ; [v] maternal parent is MRI; [u]backcross to the susceptible parent, SB22; [t]backcross to the resistant parent, MRI; [s]F_1 generations were not significantly different (P=0.335) according to Fisher's two-sample t test.

Table 2. Mean square and variance component estimates for response to downy mildew (*Peronospora belbahrii*) in the basil MRI x SB22 full-sib family pooled across northern and southern New Jersey locations in 2014.

Source	df	Mean square	Variance component
Generation	5	17.114**	-
Block	2	0.168 NS	0.000
Location	1	0.386*	0.001
Block x Location	2	0.015 NS	0.000
Generation x Location	5	0.024 NS	0.000

*Significant at the 0.05 probability level; **Significant at the 0.001; NS = not significant

Table 3. Least square means for response to downy mildew (*Peronospora belbahrii*) six generations of the basil MRI x SB22 full-sib family pooled across northern and southern New Jersey locations in 2014.

Generation	N	Mean[z]	SE
SB22[y]	30	0.967 a	0.052
MRI[x]	30	0.047 e	0.052
F_1	60	0.137 d	0.046
F_2	300	0.409 b	0.041
BC_1P1[w]	180	0.639 c	0.042
BC_1P2[v]	144	0.076 de	0.042
LSD (5%)	0.067		

[z]Values with columns followed by the same letter are not significantly different at the 0.05 *P* level according to LSD (least significant difference).
[y]susceptible parent (P1); [x]resistant parent (P2), [w]backcross to the susceptible parent; [v]backcross to the resistant parent.

Table 4. Segregation in response to downy mildew (*Peronospora belbahrii*) among six generations of the basil MRI x SB22 full-sib family pooled across northern and southern New Jersey locations in 2014.

	MRI x SB22 Family									
	Resistant	Intermediate	Susceptible	Probability of calculated $\chi 2^z$						
Generation	0 - 0.33	0.34 - 0.66	0.67 - 1.0	3:1[y]	1:2:1	9:6:1	9:7[y]	9:3:4	12:3:1	13:3[y]
SB22[x]		-	30	-	-	-	-	-	-	-
MRI[w]	30	-	-	-	-	-	-	-	-	-
F_1	60	-	-	-	-	-	-	-	-	-
F_2	155	74	71	<0.01	<0.01	<0.01	0.11	0.03	<0.01	<0.01
BC_1P1[v]	44	32	104	<0.01	<0.01	<0.01	0.04	0.63	<0.01	<0.01
BC_1P2[u]	143	1	0							

[z]Corresponding ratios for the BC_1P1 population and gene model are 1:1 (single dominant gene), 0:2:2 (single incomplete dominant gene), 1:2:1 (two additive genes), 1:3 (two complementary genes), 1:1:2 (two genes with recessive epistasis), 2:1:1 (two genes with dominant epistasis), 3:1(two genes with dominant suppression epistasis).

[y]Ratios involving two disease response classes (single dominant gene, two complementary genes, two genes with dominant suppression epistasis) correspond to the disease response segregation ratio Resistant:(Intermediate + Susceptible).

[x]susceptible parent, P1; [w]resistant parent, P2; [v]backcross to the susceptible parent; [u]backcross to the resistant

Table 5. Gene effect estimates in response to downy mildew (*Peronospora belbahrii*) in the basil MRI x SB22 full-sib family pooled across northern and southern New Jersey locations in 2014.

Gene Effect	Estimate	SE	P^z
Mean (m)	0.41	0.02	< 0.001
Additive (d)	0.56	0.02	< 0.001
Dominant (h)	-0.58	0.15	< 0.001
Additive x Additive (i)	0.88	0.14	< 0.001
Additive x Dominant (j)	0.1	0.04	<0.001
Dominant x Dominant (l)	0.06	0.22	0.769

[z]*P* values determined by t test calculation using standard errors (SE) and degrees of freedom equal to the average number of individuals within segregating generations used to calculate each gene effect

Fig 1. Phenotypic variation in the basil MRI x SB22 full sib family. A. SB22 (basil susceptible parent), B. MRI (basil resistant parent), C. F_1 (resistant), D-G. BC_1P1 siblings (backcross to basil susceptible parent), H-K. F_2 siblings.

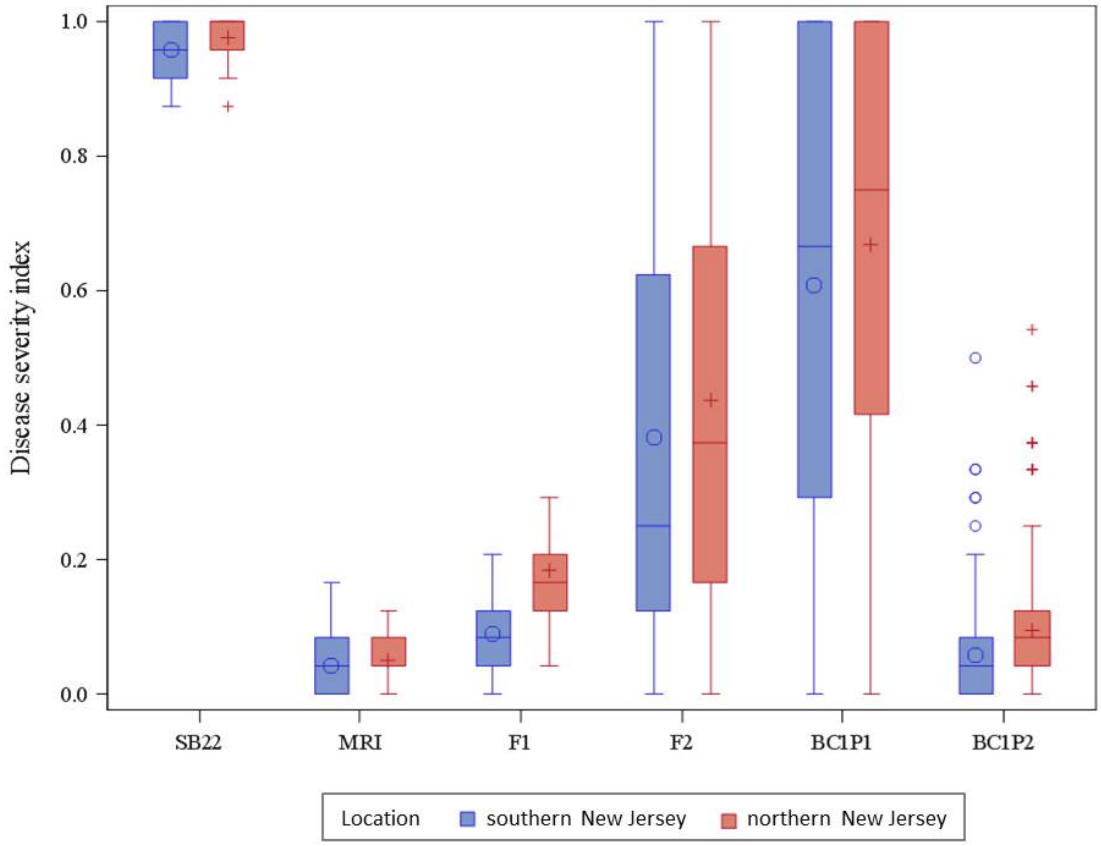


Fig 2. Downy mildew (*Peronospora belbahrii*) response distributions for basil susceptible parent (SB22), basil resistant parent (MRI), F_1, F_2, BC_1P1 (backcross to basil susceptible parent) and BC_1P2 (backcross to basil resistant parent) generations in southern (Blue, circle) and northern (Red, cross) New Jersey locations in 2014. Boxes represent generation interquartile ranges, lines within boxes represent generation medians, circles and plus signs within boxes represent generation means for southern New Jersey and northern New Jersey, respectively. Whiskers represent DSI ratings outside the interquartile range. Outliers are represented by circles and plus signs above the range of whiskers for southern New Jersey and northern New Jersey respectively.

CHAPTER 4.

A first linkage map and downy mildew resistance QTL discovery for sweet basil (*Ocimum basilicum*) facilitated by double digestion restriction site associated DNA sequencing (ddRADseq)

ABSTRACT

Limited understanding of sweet basil (*Ocimum basilicum* L.) genetics and genome structure has reduced efficiency of breeding strategies. This is evidenced by the rapid, worldwide dissemination of basil downy mildew (*Peronospora belbahrii*) in the absence of resistant cultivars. In an effort to improve available genetic resources, expressed sequence tag simple sequence repeat (EST-SSR) and single nucleotide polymorphism (SNP) markers were developed and used to genotype the MRI x SB22 F_2 mapping population, which segregates for response to downy mildew. SNP markers were generated from genomic sequences derived from double digestion restriction site associated DNA sequencing (ddRADseq). Disomic segregation was observed in both SNP and EST-SSR markers providing evidence of an *O. basilicum* allotetraploid genome structure and allowing for subsequent analysis of the mapping population as a diploid intercross. A dense linkage map was constructed using 42 EST-SSR and 1,847 SNP markers spanning 3,030.9 cM. Multiple quantitative trait loci (QTL) model (MQM) analysis identified three QTL that explained 37.6-55.1% of phenotypic variance associated with downy mildew response across three environments. A single major QTL, *dm11.1* explained 20.6-28.2% of phenotypic variance and demonstrated dominant gene action. Two minor QTL *dm9.1* and *dm14.1* explained 5.5-16.1% and 3.9-18.4% of

phenotypic variance, respectively. Evidence is provided for an additive effect between the two minor QTL and the major QTL *dm11.1* increasing downy mildew susceptibility. Results indicate that ddRADseq-facilitated SNP and SSR marker genotyping is an effective approach for mapping the sweet basil genome.

Introduction

Sweet basil (*Ocimum basilicum* L.) is the most widely cultivated and economically salient *Ocimum* species in the United States and Europe [1]. Annual US revenue generated from sweet basil and other culinary herbs sold as bedding plants are estimated to be 96.8 million dollars [2]. Introduction of basil downy mildew (*Peronospora belbahrii*) to Europe in 2001 [3] and the United States in 2007 [4] has resulted in wide spread crop destruction and an estimated tens of millions of dollars in economic losses in the US alone [5]. Absence of effective seed treatment or chemical control measures has aided rapid dissemination of this pathogen to major production areas worldwide [6–12]. Lack of economically sustainable conventional, organic or cultural control, potential for fungicide resistant pathogen evolution [6–8] and consistent disease presence in major growing regions create a compelling rationale for the development of genetic resistance to downy mildew in sweet basil.

Multiple publications have identified downy mildew resistance within the *Ocimum* genus [13,14] and Ben-Naim et al [15] demonstrated introgression into *O. x basilicum* F1 hybrids. Although possible, hybridization of commercial *O. basilicum* varieties with resistant genotypes is largely met with F1 sterility or sexual incompatibility [15]. Despite these challenges, an initial characterization of downy mildew resistance was provided by a multi-site field trial that evaluated F_2 and backcross generations derived from a cross between downy mildew resistant inbred cultivar Mrihani (MRI) and susceptible inbred Rutgers University breeding line SB22 [16]. These efforts generated important results that have helped to inform an effective breeding program targeting genetic resistance. Nevertheless, a resistant commercial sweet basil variety remains to be

seen 15 years after the first report of *P. belbahrii*, reflecting the difficulties currently associated with genetic improvement of this poorly characterized plant species.

Linkage map construction and subsequent association of DNA markers linked to important traits, or QTL, are becoming essential components of modern plant breeding programs. Although increasingly common in other horticultural species, neither a linkage map nor QTL have been developed for sweet basil. Lack of insight into the rather large [17] and potentially complex *O. basilicum* genome render sweet basil an unattractive species for genetic studies. Fortunately the rapid rise and plummeting costs of next-generation sequencing (NGS) have made high-throughput single nucleotide polymorphism (SNP) discovery accessible to non-model species [18,19]. Moreover, the introduction of reduced representation NGS strategies through restriction site associated DNA sequencing (RADseq) has revolutionized population-level genetic studies by providing a uniformly distributed subset of the genome across pooled individuals [20,21]. The combination of RADseq techniques with increasingly robust data analysis software [19,22,23] has provided unprecedented access to the genomes of complex species for linkage mapping [19]. This includes polyploid species, which can generally be divided into autopolyploid (whole genome duplication following intraspecific hybridization) or allopolyploid (whole genome duplication following interspecific hybridization) classes. Allopolyploids are more easily targeted for linkage mapping because the segregation of loci usually occurs within divergent sub-genomes, which allows for separation of homologous from homeologous loci [24,25]. By dividing segregating loci among sub-genomes one can obtain a dataset compatible with traditional diploid map construction software packages (Joinmap, RQTL, etc.) [19]. Genotyping allopolyploids is typically

performed with co-dominant markers such as simple sequence repeats (SSRs) or single nucleotide polymorphisms (SNPs) by isolation of single-locus (single sub-genome) markers [26] or development of a model for allele sub-genome assignment [27,28].

Ocimum basilicum is an outcrossing [29] tetraploid [17,30] that has demonstrated disomic inheritance for multiple traits [31,32], suggesting a diploidized polyploid genome. This allopolyploid hypothesis is supported by cytological evidence that demonstrated preferential pairing of *O. basilicum*, *O. americanum* (syn. *O. canum*) and their F_1 hybrid [16]. Furthermore, an initial investigation indicated that basil downy mildew resistance conferred genotype MRI segregates in disomic fashion [17]. Major gene control is a common form of host resistance among plant species and has been demonstrated through QTL discovery in lettuce [34–36], spinach [37], melon [38], and grape [39].

In this study, a set of genic SSR markers were developed from the currently available National Center for Biotechnology Information (NCBI) EST database and used to genotype a sweet basil F_2 mapping population from a cross between MRI and downy mildew susceptible genotype SB22 following an allotetraploid segregation model. A double digestion RADseq (ddRADseq) approach was then employed for SNP discovery and *de novo* genotyping. Genotype data was subjected to a filtration process to retain bi-allelic, homologous polymorphic loci to generate an intercross diploid dataset. Resulting genotype data were used to construct the first linkage map for sweet basil, which is anchored by SSRs and saturated by SNPs. Multiple QTL analyses were performed to identify genomic regions with association to downy mildew (*P. belbahrii*) resistance.

Materials and Methods

Plant Material

An F2 mapping population was developed in 2014 from a cross between inbred genotypes MRI (♀) and SB22 (♂) as previously described [17]. SB22 is an inbred line selected for tolerance to Fusarium wilt (*Fusarium oxysporum f. sp. basilica*) and is highly susceptible to *P. belbahrii*, while MRI is an inbred line resistant to *P. belbahrii*. The F_1 hybrid and 104 F_2 individuals were randomly selected and maintained as vegetative cuttings in Rutgers University research greenhouses (New Brunswick, NJ, U.S.A). This allowed for clones of each individual to be field transplanted for phenotyping across multiple years and locations.

Genotyping

Genomic DNA (gDNA) was extracted from the grandparents, F_1 and 104 F_2 individuals using ~80 mg of young ground leaf tissue using the E.N.Z.A. SP Plant DNA Kit (Omega BioTek, Norcross, GA). DNA was quantified and assessed for quality by measurement of 260/280 and 260/230 absorbance ratios using a Nanodrop (Thermo Scientific, Waltham, MA).

EST-SSR analysis

The National Center for Biotechnology Information (NCBI*) O. basilicum* expressed sequence tag (EST) database of 23,845 cDNA sequences was downloaded and assembled using CAP3 software [40] with default settings. The resulting contig and remaining singlet sequences were mined with SciRoKo software [41] for di-, tri- and tetranucleotide repeat sequences with a minimum of 10 nucleotides. SSRs meeting this criteria were selected for the presence of ≤300 bp flanking sequences that were subsequently used for primer pair design with Primer3 software [42]. This pipeline

produced 811 putative SSR markers from which a subset of 89 di-, 115 tri-, and 36 tetranucleotide were used in this study. Primer pairs were synthesized (Integrated DNA Technologies, Coralville, IA) with the 5' end of forward primers appended with the M13 (-21) sequence (5'-TGTAAAACGACGGCCAGT-3')[43] to facilitate fluorescent labeling of PCR products. The 5' end of reverse primers were "pig-tailed" with the 5'-GTTTCTT-3' sequence [44] to ensure consistent polyadenylation across reactions.

PCR amplification for all reactions included 5 ng of gDNA, 10x Ramp-Taq PCR buffer (Denville Scientific, Metuchen, NJ), 2.0 mM MgCl2, 0.25 mM each dNTP (Denville Scientific), 0.5 U Ramp-Taq DNA polymerase (Denville Scientific), 0.5 pmol forward primer, 1.0 pmol reverse primer, and 1.0 pmol fluorescently labeled (FAM, NED, PET, or VIC) M13(-21) primer. Template gDNA was amplified using the following conditions: initial denaturation of 94 °C for 5 min, followed by 30 cycles of 94°C for 30 sec, 55°C for 45 sec, 72 °C for 45 sec, followed by 20 cycles of 94 °C for 30 sec, 53 °C for 45 sec, 72 C for 45 sec, followed by a final extension of 72 °C for 10 min. GeneScan 600 LIZ (Applied Biosystems) size standard was added to resulting PCR products and separated by capillary electrophoresis on an ABI 3500xL Genetic Analyzer (Life Technologies Corporation, Carlsbad, CA). PCR product fragment length measurement and allele binning and was performed using Genemapper 4.1 software (Applied Biosystems).

PCR was performed in duplicate for 240 SSR markers initially using only MRI and SB22 grandparent gDNA to select for markers resulting in unambiguous PCR products (e.g., absent of any non-specific binding) and polymorphism. Markers fulfilling these criteria were subsequently used to evaluate the F_2 mapping population. SSR

markers having ≥3 amplicons for either grandparent were discarded and the remaining SSRs (containing either one or two amplicons per grandparent) were tested for chi-square goodness of fit to 1:2:1 or 3:1 diploid segregation models.

SNP Analysis

An initial experiment was performed to compare two and three enzyme library preparation approaches using gDNA from 22 F_2 individuals. The double digestion RADseq libraries were prepared according to Poland et al. [45] using rare-cutting *PstI* (NEB, USA) and the more common-cutting *MspI* (NEB, USA). In the three-enzyme digestion, the Poland et al. [45] protocol was modified to include *ApeKI*, which serves as a cutter to further reduce the complexity of the genome. Due to the addition of this enzyme, a 2 hour, 75°C incubation followed the initial digestion of gDNA with *PstI* and *MspI*. For both the two- and three-enzyme approaches, the *PstI*-complementary forward adapter and *MspI*-complementary reverse Y-adapters were added to the ligation reaction to ensure that only *PstI-MspI* fragments would be amplified during PCR, while all other digested fragments (*MspI-MspI*, *MspI-ApeKI, PstI-ApeKI* and *PstI- PstI*) should fail to amplify. Both prior to PCR and following library pooling, samples were mixed with 0.5 v/v Agencourt AMPure XP Beads (Beckman Coulter, USA) and washed with 70% ethanol to remove fragments sized less than 300 bp. All *PstI* and *MspI* adaptors included previously published barcodes [45] to uniquely identify individual samples.

All final libraries were prepared using the double digestion method rather than the triple digestion method, which generated library concentrations too low for sequencing. Four separate libraries were prepared for each grandparent to generate a 4x sequencing depth relative to the F1 and F2 progeny. Libraries for MRI (4x), SB22 (4x), F1 and 100

F2 individuals were quantified using Qubit (Life technologies, Grand island, NY). All samples were normalized to 5 ng/uL before pooling. This pool (109-plex) was paired-end sequenced on two Hi-Seq 2000 (Illumina, USA) lanes, once in Rapid-Run mode (2x150bp) and once in High-Output mode (2x100).

Sequence Data Processing

The Stacks (v1.3) software [46] pipeline was used to convert raw reads into genotype data. Concatenated single and paired end read fastq files were trimmed to 75 bp, quality filtered with the '-q' flag and de-multiplexed using the default settings of the Stacks process_radtags.pl program. The ustacks program was then used to assemble matching reads across all samples and call SNPs within each group of reads (Stack) to generate individual haplotype alleles. A minimum of 3 matching reads (-m) was required to create a Stack with a maximum nucleotide mismatch allowance (-M) of 3 for F_2 and grandparent stacks. The cstacks program generated a Catalog from Stacks that were polymorphic among the grandparents (MRI and SB22) to which F_2 progeny haplotypes were matched using the sstacks program to identify putative loci. Loci missing >20 genotyped progeny were excluded from subsequent analyses.

A potential pitfall in polyploid SNP genotyping is the vulnerability of paralogous sequence variants being called as false positive polymorphic loci [25]. In an approach similar to Hohenlohe et al [26], the Stacks web user interface was used to exclude loci with >1 SNP and significant ($p \leq 0.10$) deviation from an 1:2:1 ratio expected for an F_2 intercross between dual-homozygous grandparents. This approach was therefore employed to filter paralogs and obtain a bi-allelic SNP dataset for genetic mapping.

Linkage Map Construction

Linkage map construction was performed using Joinmap 4.1® (Kyazma, NL) [47] with genotype data coded as an F_2 intercross population type. Grouping was performed using independence logarithm of odds (LOD) scores from 2 to 15 with a step of 1, after which a minimum LOD score of 10 was used to determine autonomous linkage groups (LGs). Placement of loci was determined by comparison of map orders derived from the multipoint maximum likelihood and regression algorithms in Joinmap 4.1® (Kyazma, NL). An initial map was constructed using the maximum likelihood mapping function with parameters adjusted from the default settings when necessary to allow the algorithm to converge [48]. A second map was generated by regression mapping using a minimum LOD score of 4.0, recombination frequency of 0.35 and ripple jump threshold of 5.0. Maximum likelihood and regression maps were compared to identify suspect loci that might be misplaced. A locus or group of loci demonstrating major differences in map order location were removed to provide robust support for loci placement in the final map, which was estimated by maximum likelihood method. The maximum likelihood method assumes recombination events are independent [47] , thus recombination frequency converted to map units (cM) using Haldane's mapping function as it assumes no crossover interference.

Phenotyping

Downy mildew response for all individuals in this study was measured over two years and at two field locations by assessing the severity of abaxial leaf sporulation as described by Pyne et al [17]. Field phenotyping experiment locations were selected based consistent annual disease pressure [17,49] and susceptible check plants. Susceptible control cultivar 'DiGenova' was included in the experimental design and overhead

irrigation was applied as needed to provide uniform disease severity. Six leaves per individual were randomly sampled and scored using an ordered categorical scale of 0-4 (0=no sporulation, 1=1-10%, 2=11-25%, 3=26-50%, 4=51-100%). Individuals were assigned a value between 0 (lowest possible severity) and 1 (highest possible severity) by dividing by a maximum score of 24. Data were collected in 2014 at Northern (NJSN14) and Southern (NJRA14) New Jersey locations [17] and the Southern New Jersey location in 2015 (NJRA15). All experiments were performed in randomized complete block design with three replications. Each of the 94 genotyped F_2 individuals were phenotyped in NJRA14 and NJSN14, while 80 were phenotyped in NJRA15 due to plant death during the one year period between experiments.. Phenotypic data from a single date corresponding to the highest F_2 population mean disease severity were selected for each unique year x location combination (environment) for subsequent QTL analyses. Individuals not included in 2015 phenotyping were scored as missing data in QTL analysis.

QTL and BLASTn Analysis

A ‘forward selection’ [50] approach for identification of appropriate QTL models was implemented using the R/qtl package [51]. Single-QTL analysis using standard interval mapping was initially performed using the scanone function to detect genomic regions associated with downy mildew resistance across three environments. LOD thresholds for significance of QTL level were determined by separate permutation tests with 1,000 iterations at $\alpha = 0.05$. 1.5-LOD support intervals were calculated for all significant QTL detected. A subsequent two-dimensional (2D) analysis was performed using the scantwo function to detect QTL pairs on separate LGs. Permutation tests were

again performed with 1,000 iterations to determine significant LOD thresholds for the joint (full), conditional-interactive, interaction, additive, and conditional-additive two QTL models. An additional genome-wide scan for locus pairs within LGs was performed to account for potentially linked QTL.

Finally, a multiple-QTL model (MQM) was implemented using the fitqtl function to fit the appropriate linear model with the QTL detected from single and 2D analyses, represented as main effects. Analysis of variance (ANOVA) was performed to determine significant QTL effects and percentage of phenotypic variance explained (PVE) by each QTL. Genotype means and standard error for significant QTL were calculated and plotted using R/qtl.

The nucleotide Basic Local Alignment Search Tool (BLASTn) was implemented to query consensus sequences associated with downy mildew resistance. Reads corresponding to highest LOD score QTL and their flanking sequences were queried with a minimum cutoff of E-value < 1×10^{-5}.

Results

EST-SSR Markers

Among 240 EST-SSR markers used to genotype the grandparents, 142 primer pairs demonstrated clean (unambiguous) PCR amplification in which 1-4 unique amplicons (alleles) were classified as ‘functional’. Forty primer pairs (28.2%) (Table 1) were polymorphic and could be grouped into three bi-parental dual-homozygous genotypes: (i) one polymorphic, bi-allelic locus in one sub-genome and no locus present in the second sub-genome (Fig 1A); (ii) one monomorphic locus in one sub-genome and one polymorphic, bi-allelic locus in a second sub-genome (Fig. 1B); and (iii) one

monomorphic locus in one sub-genome and one locus with one allele present or absent (null) in a second sub-genome (Fig 1C). In scenarios (i) and (ii) the polymorphic locus was fit to an F_2 segregation ratio of 1:2:1 (a_2a_2:a_2b_2:b_2b_2) while in scenario (iii) the polymorphic locus was fit to an F_2 segregation ratio of 3:1 (b_2b_2:a_2a_2+a_2b_2).

The majority of mapped EST-SSR markers grouped in scenarios (ii) or (iii) (Fig. 1B-C) in which two loci (one monomorphic and one polymorphic) are amplified in individual sub-genomes (multi-locus markers). Six EST-SSR markers fit scenario (i) (Fig. 1A) in which a single homozygous, polymorphic locus is amplified within a single sub-genome and segregated in 1:2:1 (single locus markers). Two additional scenarios were observed for markers OBNJR2sg21 and OBNJR2cn92, in which two independently segregating loci represented putative homeologs. The former generated two allele pairs (OBNJR2sg21.1 and OBNJR2sg21.2), both segregating independently with each pair fitting a 1:2:1 ratio. The latter generated one pair (OBNJR2cn92.1) fitting 1:2:1 segregation and a second pair (OBNJR2cn92.2) exhibiting amplification in a presence (MRI), absence (SB22) fashion fitting a 3:1 segregation ratio. Thus, 42 EST-SSR markers were generated from 40 primer sets (Table 1; S1). Tests for goodness of fit provided evidence of segregation distortion for 10 ($p<0.10$) and 5 ($p<0.05$) EST-SSR markers depending on statistical stringency.

SNP Discovery and Polymorphic Loci Development

In an effort to maximize *O. basilicum* genome complexity reduction double (*PstI-MspI*) and triple (*PstI-MspI + ApeKI*) digestion library preparations were compared on a subset of the F_2 population (22 individuals). Triple digestion resulted in a mean DNA concentration of 3.4 ng/uL±1.6 considered too low for sequencing. Library preparation

without *ApeKI* (double digestion) resulted in an approximate 3-fold library concentration increase (mean=15.5ng/uL±3.63) adequate for sequencing. Thus, all libraries sequenced in this study were prepared by the double digestion method.

Illumina sequencing generated a total of 478,850,390 paired end reads with a guanine-cytosine (GC) content of 39-42%. 420,611,900 reads were retained following trimming (75bp), de-multiplexing and quality filtering. Six F_2 individual samples had a ~10x lower retained read count relative to other samples and were excluded from subsequent analysis. The average number of reads for the remaining 94 F_2 individuals was 4,021,000. The retained read count for the parents was ~4x the F_2 read count mean with 16,940,960 and 16,973,580 for MRI and SB22, respectively (S2).

Alignment of retained reads from both parents resulted in a Catalog containing 195,123 Stacks, 47,842 SNPs and 25,363 polymorphic loci. 3,492 polymorphic loci having less than 20 missing individuals could be mapped back to the parent Catalog. 2,532 polymorphic loci contained exactly 1 SNP (S3). Chi-square test of retained single-SNP loci revealed 565 with evidence for segregation distortion ($p<0.10$), or 22.3%, which were removed leaving 1,918 loci. Strict Chi-square ($p <0.10$) filtration of single SNP loci was employed to retain a bi-allelic set of loci [26]. A final set of 64 SNP markers produced identical genotypes and was removed, leaving a total of 1,954 polymorphic SNP and EST-SSR loci prior to linkage grouping.

Linkage Map Construction

Following removal of markers with poor support for grouping (LOD<10.0; rf<0.35) or placement, the final genetic map contained 1,847 SNP and 42 EST-SSR markers. The overall frequency of aa, ab and bb genotypes was 22.6%, 52.0% and 25.4%,

respectively, reflecting an F_2 dataset with adherence to Mendelian segregation. Grouping initially yielded 24 LGs as expected for the haploid chromosome set (n=24). However, two LGs exhibited large and centrally located gaps of 53.5 and 50.4 cM, resulting in unusually high total map distances of 315.2 and 310.5 cM, respectively. Multiple marker placement diagnostics did not provide evidence for any poorly supported markers causing distance inflation, thus these gaps were determined real representations and warranted the division of each LG into two: LG8 (87.1 cM), LG9 (95.6 cM) and LG14 (142.0 cM), LG15 (71.4 cM). Division of these groups had no effect on marker order. The final map yielded 26 LGs (Fig. 2) with an average distance of 166.6 cM and a total map distance of 3030.9 cM. SNP and EST-SSR markers were densely populated throughout the map and uniformly distributed among LGs, averaging a 1.6 cM distance between markers (Table 2).

Forty-two SSR markers mapped across 23 of the 26 LGs (Table 2), providing critical PCR-based "anchor" markers for potential comparison to future sweet basil linkage maps. Two primer sets resulted in two pairs of markers (OBNJR2sg21.1, OBNJR2sg21.2 and OBNJR2cn92.1, OBNJR2cn92.2) mapping to 4 LGs and may represent two homeologous chromosome sets (LGs 5,10 and 18,19) (Fig. 2).

Downy Mildew Response

Downy mildew response among individuals evaluated in 2015 was similar to previously reported results in 2014 [17]. Both grandparents were consistent in their differential response to downy mildew with SB22 >0.96 disease severity and MRI < 0.04 disease severity (Fig 3). The F_2 population mean for NJRA15 was 0.43 as compared to 0.44 and 0.39 for NJRA14 and NJSN14, respectively. The distribution across three

environments demonstrates some skewness towards a disease resistance response as previously described [17] (Fig 3). Square root transformation was applied to F_2 phenotypic data [50], resulting in normally distributed residual variance across the three environments. Thus, all subsequent QTL analyses were performed with phenotypic data square root transformed.

Detection of QTL Conferring Downy Mildew Resistance

Initial interval mapping QTL analysis detected one LG region surpassing calculated LOD thresholds ($\alpha = 0.05$) 3.99, 4.01, and 4.10 corresponding to environments NJSN14, NJRA14 and NJRA15, respectively (**Fig 4**). A maximum LOD score was associated with the most distal end of LG 11 closest to SNP marker '11636', which was renamed *dm11.1*. A 1.5 LOD confidence interval spanned approximately 45 cM from the location of *dm11.1* (**Fig 2**).

A 2D genome scan identified two additional QTL, located on LG 9 at 74.9 cM and LG 14 at 73.7 cM, renamed *dm9.1* and *dm14.1*, respectively. Interestingly, these QTL were detected ($p<0.05$) in environment NJSN14, but not in NJRA14 or NJRA15. The two-QTL model identified in NJSN14 provided evidence that *dm9.1* and *dm14.1* each act independently with *dm11.1*, thus forming two pairs: *dm9.1*, *dm11.1* and *dm11.1*, *dm14.1*. In both pairs the following two-QTL models were found to be significant: additive ($LOD \geq 8.17$; $p \leq 0.003$) and additive-conditional ($LOD \geq 4.53$; $p \leq 0.016$). These results indicate a single-QTL model is inadequate and that downy mildew resistance in at least one environment (NJSN14) should be modeled with multiple QTLs. Strong evidence ($p \leq 0.003$) was provided for the pairwise additive model suggesting that the

detected locus pairs act additively to affect response to downy mildew in the MRI x SB22 F_2 population.

Significance of two QTL pairs provided evidence for 3-QTL in at least one environment (NJSN14), necessitating the inclusion of *dm9.1*, *dm11.1* and *dm14.1* in a MQM. Results of the MQM analysis across all three environments indicated that *dm11.1* represents a 'major QTL' that consistently explained the greatest percentage of phenotypic variance (20.6 – 28.2%). The *dm14.1* QTL was well-supported (LOD=3.3; $p<0.001$) in the NJRA14 environment with 10.5% PVE, but was not detected ($p=0.11$) in single 2015 environment (NJRA15). Less support (LOD=2.1; $p<0.012$) was provided for *dm9.1* in NJRA14 and explained 6.5% of phenotypic variance (Table 3). Similarly, in NJRA15 *dm9.1* was weakly detected ($p=0.047$) with 5.5% PVE. LOD scores and phenotypic effect of these two QTL were more pronounced in environment NJSN14 where *dm9.1* (LOD=5.8; $p<0.001$) and *dm14.1* (LOD=6.5; $p<0.001$) explained 16.1% and 18.4%, respectively. Given their variable and generally lower contribution across environments *dm9.1* and *dm14.1* were considered 'minor QTL'.

Genotypes for QTL detected in the 2D genome scan were examined for their effect on downy mildew response in environment NJSN14 (Fig 5). In each QTL, 'a' alleles inherited from resistant parent MRI were associated with a lower F_2 mean downy mildew (disease) severity, while the 'b' alleles from susceptible parent SB22 were associated with a higher mean disease severity. Individuals with *dm11.1* genotypes 'aa' or 'ab' had similarly low means of 0.25±0.06 and 0.34±0.05, as compared to 0.63±0.07 for the 'bb' genotype (Table 4). Proximity of the 'ab' mean to the MRI genotype, 'aa' (Fig 5B), demonstrated dominant gene effects influence *dm11.1*-conferred downy mildew

resistance. In contrast, the homozygote (‘aa’) - heterozygote (‘ab’) - homozygote (‘bb’) trend for *dm9.1* (Fig 5A) and *dm14.1* (Fig 5C) appears relatively linear suggesting additive gene effects have greater influence over the response to downy mildew.

The genotypes of *dm9.1* (Fig 5A) and *dm14.1* (Fig 5C) were evaluated in combination with *dm11.1* genotypes and demonstrated a similar effect across interacting genotype classes. In the case of *dm9.1* in combination with *dm11.1*, dual ‘aa’ homozygotes conferred the strongest resistance (0.14±0.08) while the dual ‘bb’ homozygotes result in highest mean disease severity (0.90±0.11). When *dm11.1* is considered with *dm14.1* dual ‘aa’ and ‘bb’ homozygotes result in a mean response of 0.18±0.11 and 0.97±0.11, respectively. An intermediate phenotype was previously hypothesized for the MRI x SB22 F_2 population [17] and is supported in the presence of 3 ‘b’ alleles in both cases of QTL pairs (Fig 5D-E). Three ‘b’ alleles can be achieved with one heterozygous locus, one homozygous ‘bb’ locus and the reciprocal (genotypes at each locus switched). In the case of a heterozygous *dm11.1* and homozygous *dm14.1*, mean response is 0.59±0.07 compared to 0.69±0.09 for the reciprocal (Table 5). When considered with their associated standard errors, these two genotype scenarios appear to be associated with an intermediate response (Fig 5E). Results are similar for *dm11.1* and *dm14.1*, where the range is 0.65±0.08 - 0.66±0.11 (Table 5). Investigation of segregating alleles at multiple loci demonstrates that while *dm11.1* is most influential, it is not acting independently and a more complete informative model must include additive minor effect QTL. Both *dm9.1* and *dm14.1* separately fit a two-QTL additive model with *dm11.1*. Means for each genotype by genotype effect colored on a scale from green (low disease severity) to yellow (intermediate disease severity) to red (high disease severity).

A single consensus sequence flanking *dm11.1* at position 115.3 cM demonstrated highly significant homology (E-value < 4.0 x 10^{-17}) with a vacuolar-processing enzyme (VPE) gene predicted from *Cajanus cajan* (XM_020368305), *Glycine max* (XM_003550235) and *Fragaria vesca* (XM_004290969) mRNA. No other sequences queried exceeded the minimum cutoff E-value.

Discussion

EST-SSR and SNP Genotyping

Until recently, linkage mapping of non-model species relied heavily upon PCR-based markers such as SSRs and AFLPs, providing valuable but costly genotype data that were often inadequate for achieving dense genome coverage. Reduced representation sequencing has substantially decreased the economic and bioinformatics hurdles required to genotype and map plant species lacking genomic resources [52–54]. Little to no genomic sequence data have been made available for basil with the exception of a recent *Ocimum sanctum* draft genome assembly estimated to be 386 Mbp [55]. This massive disparity in *O. sanctum* genome size relative to *O. basilicum* suggests massive accumulation of genomic content and genetic divergence likely to exceed a threshold (1-5%) [56] at which mapping short reads would be successful. The *O. sanctum* assembly is therefore unlikely to provide utility for read alignment or validation of physical marker positions, leaving *O. basilicum* currently without a reference genome.

Implementation of *PstI-MspI* ddRADseq [45] facilitated high-throughput *de novo* SNP discovery and made feasible the construction of a first generation sweet basil linkage map. The Stacks *de novo* genotyping software pipeline performed effectively with ddRADseq sequence data, generating over 25,000 polymorphic loci between

grandparents without a reference genome. Strict filtration for single-SNP markers from dual-homozygous grandparents absent of F_2 segregation distortion ($p<0.10$) resulted in a greater than 10-fold reduction in the final number of mappable loci. However, 1,887 bi-allelic SNP markers were identified, successfully generating a dataset imputable as a diploid intercross.

The 22.3% rate of segregation distortion ($p<0.10$) observed for the MRIxSB22 F_2 population is comparable to that of better characterized allopolyploids such as strawberry (22.4%. $p < 0.05$) [57], wheat (34%) [58] and peanut (39.1%, $P<0.05$) [59]. Deviation from expected segregation ratios can be attributed to various reproductive biological factors [60,61] and are typically higher in mapping populations derived from interspecific crosses [62,63]. Phylogenetic analysis indicates the grandparents used in this study belong to the same species, *O. basilicum*, (*data unpublished*) and no sterility was observed in the F1 generation. Furthermore, preferential pairing is demonstrated by the frequent occurrence of predictable disomic inheritance patterns of EST-SSR loci (1:2:1 and 3:1), indicating divergent subgenomes that are less likely to exhibit multivalent chromosome behavior during meiosis. Instead, SNP markers failing to fit expected genotype class segregation patterns may be due to mistaken merging of sequence reads from homeologs subsequently called as homologous polymorphic loci within the same subgenome [20,25]. Availability of known diploid ancestors for the *Gossypium* sp. A_T and D_T subgenomes facilitated identification of putative SNP locus homologs [29]. In the absence of such resources, precautionary removal of loci with poor chi square goodness of fit to a 1:2:1 ratio was necessary to avoid inclusion of potential false-positive loci.

Although not critical to saturation of the linkage map, development and mapping of 42 EST-SSRs (28.1% of functional markers) (Table 1) provided needed evidence for disomic inheritance. A similar approach was recently employed to determined disomic inheritance of SSR markers in S_1 populations of *Cynodon dactylon* [64]. Eighty-seven of the 142 (61.3%) functional EST-SSR markers amplified two or more alleles, which is comparable to the 66.2% reported in a comparison inbred *Brassica* species [27]. In an inbred allotetraploid the maximum number of alleles represented by a single SSR marker in one genotype should not exceed two (Fig 1). Occurrence of 3 or 4 alleles per locus in a single grandparent (7.3% in SB22 and 8.7% in MRI) suggests a small percentage of heterozygous loci in one (3 alleles) or both (4 alleles) subgenomes. Although these SSR markers were not mappable, observation of tri- and tetra-allelic loci provide further supporting evidence for an allotetraploid genome structure.. In the absence of knowledge concerning *O. basilicum* genome structure, initial EST-SSR genotype information provided needed evidence for disomic inheritance that could be fit to a traditional diploid intercross model for further investigation.

A First Generation Sweet Basil Linkage Map

This study resulted in a sweet basil linkage map with 1,847 SNP and 42 SSR markers covering 3030.9 cM. The 26 LGs reported include two LG sets (8,9 and 14,15) that were originally merged and >300 cM in length. Observation of ~50 cM gaps in these two LGs and evidence of weak linkages among markers on either side of each gap informed the decision to divide these two LGs. This conservative approach avoided the possibility of mistakenly combining separate chromosomes and generated 26 highly supported LGs with evenly distributed markers (Table 2) and no major gaps (Fig 3). A similar result was

recently reported for cultivated strawberry from a ddRADseq-based linkage mapping generating 31 LGs, three greater than the expected 28 for the known haploid chromosome set (n=28) [54].

EST-SSR markers were successfully distributed across 23 of the 26 linkage groups in this study (Fig 1). SSR markers derived from genic sequence databases such as EST libraries are more likely to be transferrable across diverse germplasm and are thus ideal 'anchor' markers for comparison of linkage maps across populations [27,65]. Clustering of 2-4 EST-SSR markers was common on the MRI x SB22 linkage map. Among 7 LGs, SSRs (between 2 and 3) mapped within very short intervals (2.3 ± 1.7 cM) (Fig 2). The *O. basilicum* NCBI EST library is based largely on tissue-specific cDNA sequences from transcripts related to synthesis of secondary metabolites [66]. The occurrence of the EST-SSR groupings suggest they may be derived from transcripts contributing to a given biosynthetic pathway, potentially clustered in a single genomic region. Interestingly, two multi-locus SSRs (OBNJR2sg21 and OBNJR2cn92) mapped to 4 unique LGs with no evidence for segregation distortion ($p>0.10$) (Table 1). These four LGs potentially represent two homeologous chromosome sets (LGs 5,10 and 18,19), however, further investigation is needed to build support for this hypothesis.

Downy mildew resistance QTL detection

One major QTL, *dm11.1*, and two minor QTL, *dm9.1* and *dm14.1*, were associated with response to downy mildew in the MRI x SB22 F_2 mapping population. Major QTL *dm11.1* was located on the most distal end of LG 11 (160.0 cM), close to SNP '11636' and explained 20.6-28.2% of phenotypic variance across three environments. The contribution of minor QTL *dm9.1* and *dm14.1* was lower in NJRA14

and NJRA15 where the combined PVE was 17.0% and 9.4%, respectively. The 2014 F_2 mean disease severity in southern NJ (NJRA14) was significantly higher ($p<0.05$) than northern NJ (NJSN14) [17]. Despite this difference in disease pressure all three QTL were detected ($p<0.05$) in both 2014 NJ locations. However, the PVE in NJRA14 by minor QTL decreased substantially relative to NJSN14, while PVE by *dm11.1* was slightly increased by 3.3% (Table 3). Only *dm9.1* and *dm11.1* were detected in the 2015 environment NJRA15, while *dm14.1* could not be identified ($p=0.109$). PVE of 28.2% for *dm11.1* was highest in this environment, while that of *dm9.1* was comparable to NJRA14 (Table 3). Higher F_2 mean disease severities of 0.44 and 0.43 for the southern NJ location in 2014 and 2015, respectively, suggest an interaction of disease pressure and/or location with QTL effect. However, it should be noted that the population size was reduced to 80 individuals in the NJRA15, which may have affected QTL resolution in this environment. PVE of *dm11.1* was negatively correlated with *dm9.1* and *dm14.1* across all environments (Table 3), suggesting that the effect of minor and major QTL may be inversely related when subject to different environmental conditions (eg. disease severity).

When considered in isolation, this *dm11.1* would appear to act as a single-dominant gene in which one 'a' allele is sufficient to confer a resistant downy mildew response ($\leq 0.34 \pm 0.05$) (Table 5) (Fig 5B) as has been observed in *Brassica* spp. [67], spinach [37] grape [39]. However, the single, dominant gene hypothesis (0-0.33 = resistant individual and $0.34 \leq$ susceptible individual) was previously rejected by Pyne et al. [17] (Chi-squared test, $p < 0.01$) in F_2 and backcross populations using phenotypic data from NJSN14 and NJRA14, concluding that at least one additional locus was affecting

downy mildew response. In this study, detection of additional *dm9.1* and *dm14.1* in additive-two and multiple QTL models support the previous phenotypic-based findings.

Greatest support for QTL *dm9.1* and *dm14.1* (LOD> 5.8; p<0.001) occurred in environment NJSN14 (northern New Jersey; 2014) where these QTL were associated with 34.5% of PVE (Table 3). Consideration of genotype effects for both QTL in this environment demonstrates a severe consequence (high susceptibility) for individuals with dual homozygous 'b' alleles (Fig 5D-E) in which susceptibility is additive as F_2 mean disease severity surpasses the maximum (0.66±0.07) observed for any individual 'bb' QTL genotype (Fig 5D-E). This result is supported by previously identified highly significant (p<0.001) positive additive ('bb') x additive ('bb') effects from a joint scaling test using the MRI x SB22 full-sibling family indicating the presence of homozygous loci with an increase in susceptibility [17]. Successive subtraction of SB22 'b' alleles from either of the *dm11.1-dm9.1* or *dm11.1-dm14.1* QTL pairs detected in 2D QTL analysis results in an incremental reduction of F_2 mean disease severity (i.e., increased resistance). This two-QTL system therefore provides evidence for at least three downy mildew response classes: susceptible (4 'b' alleles), intermediate (3 'b' alleles) and resistant (0-2 'b' alleles) (Fig 5D-E).

Dominant gene action conferred by the 'a' allele in *dm11.1* appears to be capable of countering the susceptibility effect of 'b' alleles in either minor QTL when 1-2 'b' alleles are present. Resistance begins to break down, however, with the accumulation 3 or 4 'b' alleles in either QTL (Fig 5D-E). Again, these results are supported by a previously described positive additive ('bb') x dominant ('ab') effect [17] through resistance of the heterozygous *dm11.1* locus being reduced by homozygous 'bb' loci in either minor QTL.

The resistance-reducing effect of a ‘bb’ genotype in *dm11.1* with an ‘aa’ genotype in *dm9.1* was greater than the reciprocal (‘aa’ genotype in *dm11.1* with ‘bb’ *dm9.1*) (Fig 5D). In contrast, comparison of F_2 means for reciprocal, opposing homozygous genotype combination in *dm11.1* and *dm14.1* resulted in similar mean disease severity (Fig 5E). In both cases, a relatively high SE (0.07-0.10) demonstrates variability in downy mildew response when the recessive ‘bb’ *dm11.1* genotype is present with the ‘aa’ genotype of either minor QTL, resulting in some loss of resistance response within the population. Chi-square goodness of fit to complementary and recessive epistatic gene models in F_2 and backcross generations from phenotypic data suggested dominant effects were needed to confer resistance [17]. A dominant (‘ab’) x dominant (‘ab’) gene effect was thus expected but unsupported (p=0.769)[17]. QTL in this study provide evidence for a more complex gene model with a major, single dominant and two minor, additive QTL (Table 3).

It is clear that applied resistance breeding would benefit from ensuring germplasm have the *dm11.1* ‘aa’ (MRI) genotype as a heterozygous locus at *dm11.1* will result in loss of resistance through segregation during self-pollinated seed propagation. Given the potential increased susceptibility effect of the ‘b’ allele, its removal from each QTL (‘aa’ genotype) would be preferable to ensure a high level of stable resistance. A similar model was identified in the GR x Ice RIL population in which the homozygous Iceberg genotype at two QTL conferred significantly higher resistance to downy mildew [68].

Quantitative and qualitative forms of downy mildew resistance have been reported in multiple plant species such as *Cucumis* spp. where sources of host resistance have been identified and characterized for decades [69]. Quantitative resistance, while

less susceptible to breakdown, is subject to greater variation [69] across environments as was observed for the QTL *dm9.1* and *dm14.1* detected in this study (9.4-34.5 % PVE). A similar range of PVE (15-30%) for downy mildew response was reported for two additive QTL in cucumber $F_{2:3}$ families across three environments [38]. Qualitative downy mildew resistance, while more prone to breakdown, is less vulnerable to environmental interaction and often associated with gene-for-gene host-pathogen interaction. Major QTL *dm11.1* was detected in a distal region of LG 11 with resistance conferred from grandparent MRI by the dominant 'a' allele. The downy mildew resistance dominant locus *Rpv3* was also detected in a distal chromosomal region of *V. vinifera* 'Bianca' known to contain NBS-LRR gene clusters [39]. Development of a *de novo* metatranscriptomics pipeline [70] for *O. basilicum* provides a platform for identification of resistant gene motifs, which could potentially be mapped to genomic regions containing QTL such as *dm11.1* for functional characterization.

A majority of QTL and flanking-QTL consensus sequences did not demonstrate significant homology with the NCBI nucleotide database indicating that these markers are unlikely to be in genic regions. However, a single consensus sequence flanking *dm11.1* exhibited a significant BLAST hit with predicted gene sequences for the cysteine proteinase VPE gene. Activity of VPE has been shown to play a role in programmed cell death as part of a hypersensitive response triggered by oomycete elicitors [71]. Increased VPE activity has also been positively associated with *H. arabidopsis* sporulation, independent of programmed cell death, during host hypersensitive response [72]. Although the specific role of VPE in the *O. baslicum* - *P. belbahrii* pathosystem cannot

be deduced, ubiquity of this enzyme in plant defense response [73] warrants further investigation.

RADseq approaches have changed the landscape of linkage and QTL mapping for non-model plant species by introducing low-cost, high-throughput, *de novo* SNP discovery. The inexpensive acquisition of tens or hundreds of thousands of SNP markers allow researchers working with poorly understood genomes to be 'picky', applying strict filtration to retain >1,000 high-quality SNPs with predictable segregation patterns. This reduces the burden to generate large numbers of labor-intensive, costly markers such as SSRs, instead utilizing a smaller subset to serve as 'anchor' markers for subsequent map comparison. This genotyping approach high-SNP/low-SSR genotyping approach has facilitated map construction in peach [53], strawberry [54], sesame [74] and lentil [75]. In this study, the power of this approach is further demonstrated through the development of a sweet basil linkage map and QTL detection.

Conclusions

Genetic study of non-model, horticultural species such as sweet basil are often neglected due to perceived low economic importance; however, this renders such crops vulnerable to rapid and wide-spread decline upon introduction of new plant pathogens. *P. belbahrii* now causes worldwide economic losses with no available resistant sweet basil cultivars. In this study, a set of EST-SSR markers were developed and mapped in the MRI x SB22 F_2 sweet basil mapping population providing molecular evidence of disomic inheritance. Effective filtration of ddRADseq SNP markers generated 1,847 bi-allelic, polymorphic markers in the absence of a reference genome. This novel genotyping approach facilitated construction of the first linkage map for sweet basil. The utility of

this map was demonstrated through identification of one major and two minor QTL associated with downy mildew resistance, largely supporting a previous report using phenotypic data only. Results provide the first steps towards the development of molecular tools for accelerated sweet basil breeding strategies.

References

1. Simon, J.E, Quinn, J, Murray JG. Basil: A Source of Essential Oils. In: Janick, J.; Simon JE, editor. Advances in New Crops. Portland, OR: Timber Press; 1990. pp. 484–489.
2. USDA. United States Department of Agriculture Census of Agriculture. https://www.agcensus.usda.gov. (Accessed October 1, 2016).
3. Belbahri L, Calmin G, Pawlowski J, Lefort F. Phylogenetic analysis and real time PCR detection of a presumbably undescribed *Peronospora* species on sweet basil and sage. Mycol Res. 2005;109: 1276–1287. doi:10.1017/S0953756205003928
4. Roberts PD, Raid RN, Harmon PF, Jordan SA, Palmateer AJ. First Report of Downy Mildew Caused by a *Peronospora* sp. on Basil in Florida and the United States. Plant Dis. 2009;93: 199. doi:10.1094/PDIS-93-2-0199B
5. Wyenandt CA, Simon JE, Pyne RM, Homa K, McGrath MT, Zhang S, et al. Basil Downy Mildew (*Peronospora belbahrii*): Discoveries and Challenges Relative to Its Control. Phytopathology. 2015;105: 885–94. doi:10.1094/PHYTO-02-15-0032-FI
6. Blomquist CL, Rooney-Latham S, Nolan PA. First Report of Downy Mildew on Field-Grown Sweet Basil Caused by a *Peronospora* sp. in San Diego County, California. Plant Dis. 2009;93: 968. doi:10.1094/PDIS-93-9-0968A
7. Garibaldi A, Minuto A, Minuto G, Gullino ML. First Report of Downy Mildew on Basil (*Ocimum basilicum*) in Italy. Plant Dis. 2004;88: 312. doi:10.1094/PDIS.2004.88.3.312A
8. Kanetis L, Vasiliou A, Neophytou G, Samouel S, Tsaltas D. First Report of Downy Mildew Caused by *Peronospora belbahrii* on Sweet Basil (*Ocimum basilicum*) in Cyprus. Plant Dis. 2013;98: 283. doi:10.1094/PDIS-07-13-0759-PDN
9. McLeod A, Coertze S, Mostert L. First Report of a *Peronospora* Species on Sweet Basil in South Africa. Plant Dis. 2006;90: 1115. doi:10.1094/PD-90-1115A
10. Nagy G, Horváth A. Occurrence of Downy Mildew Caused by *Peronospora belbahrii* on Sweet Basil in Hungary. Plant Dis. 2011;95: 1034. doi:10.1094/PDIS-04-11-0329
11. Saude C, Westerveld S, Filotas M, McDonald MR. First Report of Downy Mildew Caused by Peronospora belbahrii on Basil (*Ocimum* spp.) in Ontario. Plant Dis. 2013;97: 1248. doi:10.1094/PDIS-01-13-0026-PDN
12. Kong XY, Wang S, Wan SL, Xiao CL, Luo F, Liu Y. First Report of Downy Mildew on Basil (*Ocimum basilicum*) in China. Plant Dis. 2015;99: 1642. doi:10.1094/PDIS-01-15-0077-PDN
13. Pyne RM, Koroch AR, Wyenandt CA, Simon JE. A Rapid Screening Approach to Identify Resistance to Basil Downy Mildew (*Peronospora belbahrii*). HortScience. 2014;49: 1041–1045.
14. Djalali Farahani-Kofoet R, Römer P, Grosch R. Systemic spread of downy mildew in basil plants and detection of the pathogen in seed and plant samples. Mycol Prog. 2012;11: 961–966. doi:10.1007/s11557-012-0816-z
15. Ben-Naim, Y, Falach, L, Cohen Y. Resistance against basil downy mildew in

Ocimum species. Phytopathology. 2015;105: 778-785. 2710. doi:10.1094/PHYTO-11-14-0295-R

16. Pushpangadan, P. and Sobti, S.N. Cytogenetical studies in the genus *Ocimum*. I. Origin of *O. americnum*, cytotaxonomical and experimental proof. Cytologia. 1982. 47: 575-583.
17. Pyne RM, Koroch AR, Wyenandt CA, Simon JE. Inheritance of Resistance to Downy Mildew in Sweet Basil. J Am Soc Hortic Sci. 2015;140: 396–403.
18. Koroch AR, Wang W, Michael TP, Dudai N, Simon JE, Belanger FC. Estimation of nuclear DNA content of cultivated *Ocimum* species by using flow cytometry. Isr J Plant Sci. 2010;58: 183–189. doi:10.1560/IJPS.59.3-4.183
19. Davey JW., Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12: 499–510. doi:10.1038/nrg3012
20. Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA. Single nucleotide polymorphism identification in polyploids: A review, example, and recommendations. Mol Plant. Elsevier Ltd; 2015;8: 831–846. doi:10.1016/j.molp.2015.02.002
21. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3: e3376. doi:10.1371/journal.pone.0003376
22. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011;6: 1–10. doi:10.1371/journal.pone.0019379
23. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: An analysis tool set for population genomics. Mol Ecol. 2013;22: 3124–3140. doi:10.1111/mec.12354
24. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One. 2014;9. doi:10.1371/journal.pone.0090346
25. Christensen K a, Brunelli JP, Lambert MJ, DeKoning J, Phillips RB, Thorgaard GH. Identification of single nucleotide polymorphisms from the transcriptome of an organism with a whole genome duplication. BMC Bioinformatics. 2013;14: 325. doi:10.1186/1471-2105-14-325
26. Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol Ecol Resour. 2011;11: 117–122. doi:10.1111/j.1755-0998.2010.02967.x
27. Li H, Younas M, Wang X, Li X, Chen L, Zhao B, et al. Development of a core set of single-locus SSR markers for allotetraploid rapeseed (*Brassica napus* L.). Theor Appl Genet. 2013;126: 937–947. doi:10.1007/s00122-012-2027-z
28. van Dijk T, Noordijk Y, Dubos T, Bink MC, Meulenbroek BJ, Visser RG, et al. Microsatellite allele dose and configuration establishment (MADCE): an integrated approach for genetic studies in allopolyploids. BMC Plant Biol. 2012;12: 25. doi:10.1186/1471-2229-12-25
29. Logan-Young CJ, Yu JZ, Verma SK, Percy RG, Pepper AE. SNP Discovery in Complex Allotetraploid Genomes (*Gossypium* spp., Malvaceae) Using

Genotyping by Sequencing. Appl Plant Sci. 2015;3: 1400077. doi:10.3732/apps.1400077

30. Nation RG, Janick J, Simon JE. Estimation of outcrossing in basil. Hortic Sci. 1992;27: 1221–1222.
31. Khosla MK, Tawi- J, Group S. Karyomorphological Studies in Genus *Ocimum*. Cytologia. 1985;50: 253–263.
32. Phippen WB, Simon JE. Anthocyanin inheritance and instability in purple basil (Ocimum basilicum L.). J Hered. 2000;91: 289–296. doi:10.1093/jhered/91.4.289
33. Chaimovitsh, D.;Dudai, N., Putievesky, E.; Ashri A. Inheritance of Resistance to Fusarium Wilt in Sweet Basil. Phytopathology. 1983;90: 58–60. doi:10.2135/cropsci1983.0011183X002300010010x
34. Jeuken M, Lindhout P. Lactuca saligna, a non-host for lettuce downy mildew (*Bremia lactucae*), harbors a new race-specific *Dm* gene and three QTLs for resistance. Theor Appl Genet. 2002;105: 384–391. doi:10.1007/s00122-002-0943-z
35. Zhang NW, Lindhout P, Niks RE, Jeuken MJW. Genetic dissection of *Lactuca saligna* nonhost resistance to downy mildew at various lettuce developmental stages. Plant Pathol. 2009;58: 923–932. doi:10.1111/j.1365-3059.2009.02066.x
36. den Boer E, Pelgrom KTB, Zhang NW, Visser RGF, Niks RE, Jeuken MJW. Effects of stacked quantitative resistances to downy mildew in lettuce do not simply add up. Theor Appl Genet. 2014;127: 1805–1816. doi:10.1007/s00122-014-2342-7
37. Irish BM, Correll JC, Feng C, Bentley T, de Los Reyes BG. Characterization of a resistance locus (*Pfs-1*) to the spinach downy mildew pathogen (*Peronospora farinosa* f. sp. *spinaciae*) and development of a molecular marker linked to *Pfs-1*. Phytopathology. 2008;98: 894–900. doi:10.1094/PHYTO-98-8-0894
38. Wang Y, VandenLangenberg K, Wehner TC, Kraan PAG, Suelmann J, Zheng X, et al. QTL mapping for downy mildew resistance in cucumber inbred line WI7120 (PI 330628). Theor Appl Genet. 2016; doi:10.1007/s00122-016-2719-x
39. Bellin D, Peressotti E, Merdinoglu D, Wiedemann-Merdinoglu S, Adam-Blondon AF, Cipriani G, et al. Resistance to *Plasmopara viticola* in grapevine "Bianca" is controlled by a major dominant gene causing localised necrosis at the infection site. Theor Appl Genet. 2009;120: 163–176. doi:10.1007/s00122-009-1167-2
40. Huang X, Madan a. CAP 3: A DNA sequence assembly program. Genome Res. 1999;9: 868–877. doi:10.1101/gr.9.9.868
41. Kofler R, Schlötterer C, Lelley T. SciRoKo: A new tool for whole genome microsatellite search and investigation. Bioinformatics. 2007;23: 1683–1685. doi:10.1093/bioinformatics/btm157
42. Rozen S, Skaletksy H. Primer 3 on the WWW for general uses and for biologist programmers. Bioinforma Methods Protoc Methods Mol Biol. 2000; 365–386.
43. Schuelke M. An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol. 2000;18: 233–234. doi:10.1038/72708
44. Brownstein MJ, Carpten JD, Smith JR. Modulation of non-templated nucleotide addition by Taq DNA polymerase: Primer modifications that facilitate genotyping. Biotechniques. 1996;20: 1004–1010. doi:10.2144/000113156
45. Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density

genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One. 2012;7. doi:10.1371/journal.pone.0032253

46. Catchen JM, Amores A., Hohenlohe P, Cresko W, Postlethwait JH, De Koning D-J. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. Genes|Genomes|Genetics. 2011;1: 171–182. doi:10.1534/g3.111.000240
47. Van Ooijen JW. JoinMap ® 4. JoinMap. 2006; Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.
48. Jansen J, De Jong AG, Van Ooijen JW. Constructing dense genetic linkage maps. Theor Appl Genet. 2001;102: 1113–1122. doi:10.1007/s001220000489
49. Wyenandt CA, Simon JE, McGrath MT, Ward DL. Susceptibility of basil cultivars and breeding lines to downy mildew (*Peronospora belbahrii*). HortScience. 2010;45: 1416–1419.
50. Broman KW. Review of statistical methods for QTL mapping in experimental crosses. Lab Anim (NY). 2001;30: 44–52. doi:11469113
51. Broman KW, Wu H, Sen Ś, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 2003;19: 889–890. doi:10.1093/bioinformatics/btg112
52. Henning JA, Gent DH, Twomey MC, Townsend MS, Pitra NJ, Matthews PD. Genotyping-by-sequencing of a bi-parental mapping population segregating for downy mildew resistance in hop (*Humulus lupulus* L.). Euphytica. 2016;208: 545–559. doi:10.1007/s10681-015-1600-3
53. Bielenberg DG, Rauh B, Fan S, Gasic K, Abbott AG, Reighard GL, et al. Genotyping by sequencing for SNP-based linkage map construction and QTL analysis of chilling requirement and bloom date in peach (*Prunus persica* (L.) Batsch). PLoS One. 2015;10: 1–14. doi:10.1371/journal.pone.0139406
54. Davik J, Sargent DJ, Brurberg MB, Lien S, Kent M, Alsheikh M. A ddRAD based linkage map of the cultivated Strawberry, *Fragaria x ananassa*. PLoS One. 2015;10: 1–10. doi:10.1371/journal.pone.0137746
55. Rastogi S, Kalra A, Gupta V, Khan F, Lal RK, Tripathi AK, et al. Unravelling the genome of Holy basil: an “incomparable” “elixir of life” of traditional Indian medicine. BMC Genomics. 2015;16: 413. doi:10.1186/s12864-015-1640-z
56. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 2012;7. doi:10.1371/journal.pone.0037135
57. Isobe SN, Hirakawa H, Sato S, Maeda F, Ishikawa M, Mori T, et al. Construction of an integrated high density simple sequence repeat linkage map in cultivated strawberry (*Fragaria x ananassa*) and its applicability. DNA Res. 2013;20: 79–92. doi:10.1093/dnares/dss035
58. Alheit K V, Reif JC, Maurer HP, Hahn V, Weissmann E a, Miedaner T, et al. Detection of segregation distortion loci in triticale (x *Triticosecale* Wittmack) based on a high-density DArT marker consensus genetic linkage map. BMC Genomics. 2011;12: 380. doi:10.1186/1471-2164-12-380
59. Zhou X, Xia Y, Ren X, Chen Y, Huang L, Huang S, et al. Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA

sequencing (ddRADseq). BMC Genomics. 2014;15: 351. doi:10.1186/1471-2164-15-351

60. Fishman L, Willis JH. A novel meiotic drive locus almost completely distorts segregation in *Mimulus* (monkeyflower) hybrids. Genetics. 2005;169: 347–353. doi:10.1534/genetics.104.032789
61. Taylor DR, Ingvarsson PK. Common features of segregation distortion in plants and animals. Genetica. 2003;117: 27–35. doi:10.1023/A:1022308414864
62. Manrique-Carpintero NC, Coombs JJ, Veilleux RE, Buell CR, Douches DS. Comparative Analysis of Regions with Distorted Segregation in Three Diploid Populations of Potato. G3 Genes|Genomes|Genetics. 2016;6: 2617–2628. doi:10.1534/G3.116.030031
63. Reflinur, Kim B, Jang SM, Chu S-H, Bordiya Y, Akter MB, et al. Analysis of segregation distortion and its relationship to hybrid barriers in rice. Rice. 2014;7: 3. doi:10.1186/s12284-014-0003-8
64. Guo Y, Wu Y, Anderson JA, Moss JQ, Zhu L. Disomic inheritance and segregation distortion of SSR markers in two populations of *Cynodon dactylon* (L.) Pers. var. dactylon. PLoS One. 2015;10: 1–10. doi:10.1371/journal.pone.0136332
65. Gonzalo MJ, Oliver M, Garcia-Mas J, Monfort A, Dolcet-Sanjuan R, Katzir N, et al. Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.). Theor Appl Genet. 2005;110: 802–811. doi:10.1007/s00122-004-1814-6
66. Gang DR, Wang J, Dudareva N, Nam KH, Simon JE, Lewinsohn E, et al. An Investigation of the Storage and Biosynthesis of Phenylpropenes in Sweet Basil. Plant Physiol. 2001;125: 539–555. doi:10.1104/pp.125.2.539
67. Vicente JG, Gunn ND, Bailey L, Pink DAC, Holub EB. Genetics of resistance to downy mildew in Brassica oleracea and breeding towards durable disease control for UK vegetable production. Plant Pathol. 2012;61: 600–609. doi:10.1111/j.1365-3059.2011.02539.x
68. Simko I, Atallah AJ, Ochoa OE, Antonise R, Galeano CH, Truco MJ, et al. Identification of QTLs conferring resistance to downy mildew in legacy cultivars of lettuce. Sci Rep. 2013;3: 2875. doi:10.1038/srep02875
69. Olczak-Woltman H, Marcinkowska J, Niemirowicz-Szczytt K. The genetic basis of resistance to downy mildew in *Cucumis* spp.-latest developments and prospects. J Appl Genet. 2011;52: 249–255. doi:10.1007/s13353-011-0030-8
70. Guo L, Allen KS, Deiulio GA, Zhang Y, Madeiras AM, Wick RL, et al. A de-novo-assembly-based Data Analysis Pipeline for Plant Obligate Parasite Metatranscriptomic Studies. Front Plant Sci. 2016;7: 925. doi:10.3389/FPLS.2016.00925
71. Zhang H, Dong S, Wang M, Wang W, Song W, Dou X, et al. The role of vacuolar processing enzyme (VPE) from Nicotiana benthamiana in the elicitor-triggered hypersensitive response and stomatal closure. J Exp Bot. 2010;61:3799-3812. Doi:10.1093/jxb/erq189
72. Misas-Villamil JC, Toenges G, Kolodziejek I, Sadaghiani AM, Kaschani F, Colby T, et al. Activity profiling of vacuolar processing enzymes reveals a role for VPE during oomycete infection. Plant J. 2013;73: 689-700. doi:10.1111/tpj.12062

73. Hatsugai N, Yamada K, Goto-Yamada S, Hara-Nishimura I. Vacuolar processing enzyme in plant programmed cell death. Front Plant Sci. 2015;6: 234.doi:10.3389/fpls.2015.00234
74. Wu K, Liu H, Yang M, Tao Y, Ma H, Wu W, et al. High-density genetic map construction and QTLs analysis of grain yield-related traits in Sesame (*Sesamum indicum* L.) based on RAD-Seq techonology. BMC Plant Biol. 2014;14: 274. doi:10.1186/s12870-014-0274-7
75. Ates D, Sever T, Aldemir S, Yagmur B, Temel HY, Kaya HB, et al. Identification QTLs controlling genes for se uptake in Lentil seeds. PLoS One. 2016;11.

Table 1. Description of 42 mapped EST-SSR markers in the MRIxSB22 linkage map indicating nucleotide sequence source, repeat motif, linkage group, centimorgan position and chi-square goodness-of-fit test results

Marker ID[a]	Source[b]	Motif[c]	LG	Position (cM)	Ratio	$\chi 2$	P^d
OBNJR3cn391	Contig3280	(TCA)5	1	16.406	3:1	3.33	<0.10
OBNJR2cn104	Contig3401	(CT)14	1	18.956	3:1	0.24	-
OBNJR3sg98	DY337354	(CCT)8	1	19.164	3:1	2.97	<0.10
OBNJR3cn201	Contig1783	(GAA)6	1	59.78	1:2:1	3.34	-
OBNJR2sg34	DY331634	(CT)11	2	32.395	1:2:1	0.77	-
OBNJR2cn79	Contig2573	(AT)9	2	32.665	3:1	0.52	-
OBNJR3cn362	Contig2969	(TGA)6	3	78.015	1:2:1	1.47	-
OBNJR3cn56	Contig582	(AGG)7	3	84.661	1:2:1	5.16	<0.10
OBNJR2cn80	Contig2575	(TA)12	4	90.347	1:2:1	1.68	-
OBNJR2sg21.1	DY339566	(TC)13	5	33.87	1:2:1	2.67	-
OBNJR3sg19	DY343509	(TCA)6	6	18.9	3:1	3.49	<0.10
OBNJR2sg15	DY340778	(GA)25	6	50.811	1:2:1	2.25	-
OBNJR4sg06	DY338242	(ACAA)5	7	139.338	3:1	1.65	-
OBNJR2cn78	Contig2475	(AT)10	8	56.362	1:2:1	1.88	-
OBNJR2sg21.2	DY339566	(TC)13	10	65.708	1:2:1	1.62	-
OBNJR3cn389	Contig3254	(GCA)8	11	132.024	3:1	0.86	-
OBNJR4cn11	Contig1679	(TCAC)4	12	143.958	3:1	0.06	-
OBNJR2cn83	Contig2631	(GA)18	12	144.364	1:2:1	2.65	-
OBNJR2cn17	Contig606	(AT)10	12	146.577	3:1	0.23	-
OBNJR2sg04	DY343638	(GA)17	13	40.159	1:2:1	5.37	<0.10
OBNJR4cn16	Contig2294	(CAAA)4	13	42.926	1:2:1	3.61	-
OBNJR2sg119	DY333250	(GTA)7	13	49.715	1:2:1	10.14	<0.01
OBNJR2cn29	Contig1138	(AC)16	13	54.583	1:2:1	2.25	-
OBNJR2sg33	DY333933	(AC)16	14	82.696	1:2:1	0.03	-
OBNJR4cn15	Contig2242	(GCCT)5	15	0	1:2:1	7.13	<0.05
OBNJR3cn358	Contig2910	(TCC)7	15	24.042	1:2:1	2.31	-
OBNJR3cn356	Contig2907	(AAG)9	15	26.775	1:2:1	0.48	-
OBNJR2sg31	DY336298	(AG)9	16	106.934	1:2:1	0.07	-
OBNJR3cn328	Contig2750	(AAG)6	16	107.525	1:2:1	0.76	-
OBNJR3cn192	Contig1724	(ACA)8	16	161.484	3:1	0.13	-
OBNJR3cn243	Contig2153	(GAA)6	17	0	3:1	0.06	-
OBNJR3sg177	DY322989	(TGC)5	17	2.531	3:1	4.28	<0.01
OBNJR3cn401	Contig3437	(CAG)7	17	6.377	3:1	0.97	-
OBNJR3cn377	Contig3126	(TAT)6	18	30.899	1:2:1	1.8	-
OBNJR2cn92.2	Contig3041	(CT)13	18	78.067	3:1	0.84	-
OBNJR3cn54	Contig573	(TTA)18	19	88.669	1:2:1	0.4	-
OBNJR2cn92.1	Contig3041	(CT)13	19	89.277	1:2:1	0.64	-
OBNJR3cn217	Contig1936	(ATT)6	20	32.066	3:1	0.37	-

OBNJR4sg01	DY343743	(TCCC)5	21	91.928	1:2:1	1.27	-
OBNJR3cn239	Contig2134	(TTC)8	24	58.002	1:2:1	1.2	-
OBNJR2cn38	Contig1352	(CA)13	25	55.999	1:2:1	1.2	-
OBNJR2cn73	Contig2250	(CT)24	26	56.528	1:2:1	7.13	<0.05

[a]SSR markers failing to map or found to be located at identical positions to other SSR markers are excluded. A single primer set resulting in two independently segregating loci are indicated by marker name followed by either '.1' or '.2'

[b]SSRs are sourced from CAP3-assembled NCBI *O. basilicum* EST sequence database. NCBI genbank nucleotide accession is provided for SSRs located in EST sequences that could not be assembled into contigs.

[c]Repeat motif sequence and number reported refer to the original Genbank (parent) sequences

[d]Chi-square goodness-of-fit tests resulting in $p < 0.10$ indicate evidence for segregation distortion.

Table 2. Summary of the MRIxSB22 F_2 linkage map including number of SNPs and EST-SSRs, centimorgan length and average centimorgan distance between markers for each linkage group

LG	SNPs	SSRs	Distance (cM)	Average distance between markers (cM)
1	158	4	199.3	1.2
2	68	2	69.1	1
3	48	2	93.9	1.9
4	90	1	112.3	1.2
5	71	1	121.3	1.7
6	37	2	149	3.9
7	141	1	198.7	1.4
8	52	1	87.1	1.7
9	45	0	95.6	2.2
10	87	1	123.8	1.4
11	85	1	160	1.9
12	143	3	156.9	1.1
13	57	4	89.5	1.5
14	97	1	142.9	1.5
15	37	3	71.4	1.8
16	105	3	161.5	1.5
17	91	3	126.8	1.4
18	91	2	110.3	1.2
19	65	2	122.6	1.9
20	60	1	113.7	1.9
21	57	1	110.6	1.9
22	40	0	88.8	2.3
23	34	0	86.3	2.6
24	31	1	79.2	2.6
25	37	1	103.8	2.8
26	20	1	56.5	2.8
Overall	1847[a]	42[a]	3030.9[a]	1.6[b]

[a]Total

[b]Average centimorgan distance

Table 3. Summary of three downy mildew resistance QTL detected using a multiple QTL model (MQM) across three environments.

QTL	LG	Position (cM)	SNP[a]	Confidence Interval (cM)[b]	Environment	LOD	*P*[c]	PVE (%)[d]
dm9.1	9	74.9	95799	51.9 - 95.6	NJSN14	5.8	<0.001	16.1
		-	-	-	NJRA14	2.1	0.012	6.5
		-	-	-	NJRA15	1.5	0.047	5.5
dm11.1	11	160.0	11636	115.3 - 160.0	NJSN14	7.2	<0.001	20.6
		160.0	11636	115.3 - 160.0	NJRA14	6.7	<0.001	23.3
		160.0	11636	114.0 - 160.0	NJRA15	6.5	<0.001	28.2
dm14.1	14	73.7	120555	65.3 - 92.1	NJSN14	6.6	<0.001	18.4
		73.7	120555	65.3 - 131.0	NJRA14	3.3	<0.001	10.5
		-	-	-	NJRA15	1.1	0.109	3.9

[a]Single nucleotide polymorphism (SNP) marker located in closest proximity to the QTL location
[b]1.5 LOD score intervals shown or significant (P<0.01) QTL only
[c]*P*-values represent the significance of LOD scores determined by permutation tests with 1,000 iterations at $\alpha = 0.05$
[d]Percent phenotypic variance explained

Table 4. F_2 means for downy mildew (disease) response in environment NJSN14 according to QTL genotype.

QTL	aa	ab	bb
dm9.1	0.21±0.08	0.37±0.05	0.54±0.06
dm11.1	0.25±0.06	0.34±0.05	0.66±0.07
dm14.1	0.21±0.07	0.35±0.04	0.63±0.06

Table 5. F_2 means for downy mildew (disease) response in environment NJSN14 according to two-QTL genotype by genotype combinations.

	dm9.1			***dm14.1***		
dm11.1	aa	ab	bb	aa	ab	bb
aa	0.14±0.12	0.19±0.11	0.30±0.07	0.18±0.08	0.20±0.08	0.38±0.10
ab	0.18±0.10	0.27±0.05	0.67±0.11	0.16±0.11	0.22±0.07	0.60±0.07
bb	0.35±0.15	0.654±0.08	0.90±0.11	0.30±0.11	0.69±0.09	0.98±0.11

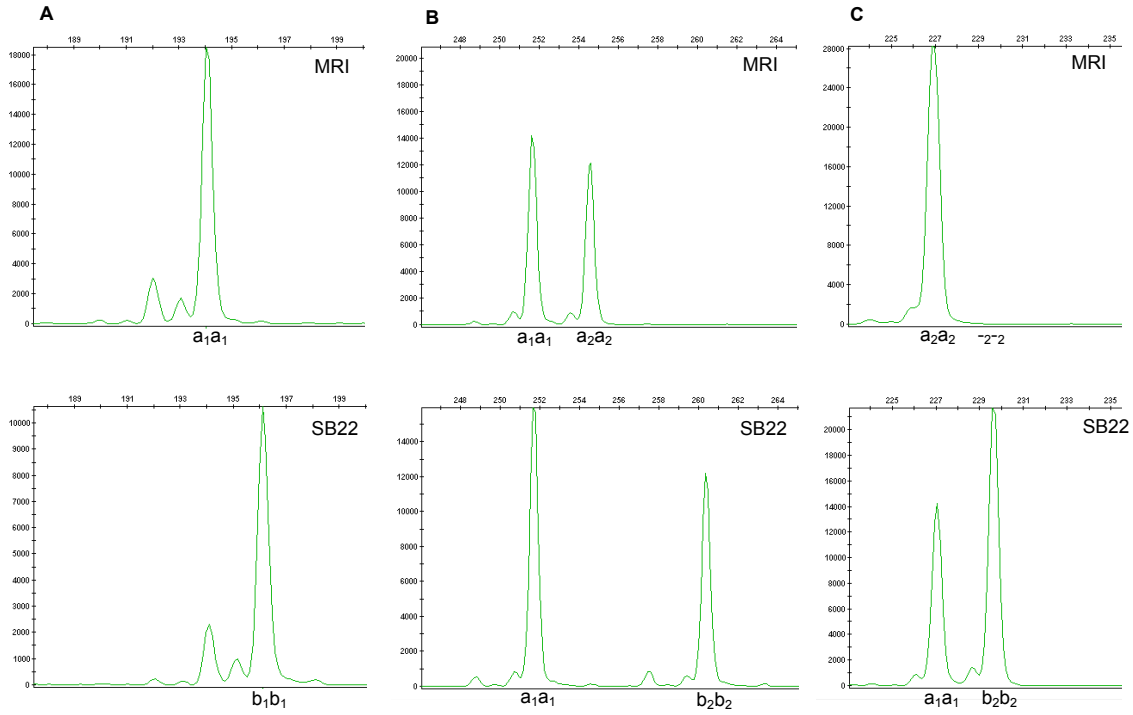


Fig 1. Polymorphic EST-SSR genotypes observed among subgenomes of homozygous grandparent genotypes MRI and SB22. Electropherogram plots for alleles represented as peaks with size (nucleotides) represented along the x-axis. (A) Marker OBNJR2sg34 genotype: One polymorphic locus within a single subgenome (arbitrarily designated with subscript 1) corresponding to MRI genotype a_2a_2 and SB22 genotype b_2b_2 with expected F_2 segregation ratio 1:2:1 (a_2a_2:a_2b_2:b_2b_2). (B) Marker OBNJR3cn328 genotype: One monomorphic locus corresponding to a single sub-genome genotype a_1a_1 and one polymorphic locus in another sub-genome (arbitrarily designated with subscript 2) with expected F_2 segregation ratio 1:2:1 (a_2a_2:a_2b_2:b_2b_2). (C) Marker OBNJR3cn80 genotype: Monomorphic locus a_1a_1 and polymorphic locus in another sub-genome corresponding to MRI null genotype $-_2-_2$ and SB22 genotype b_2b_2 with expected F_2 segregation ratio 3:1 ($-_2-_2$: b_2b_2).

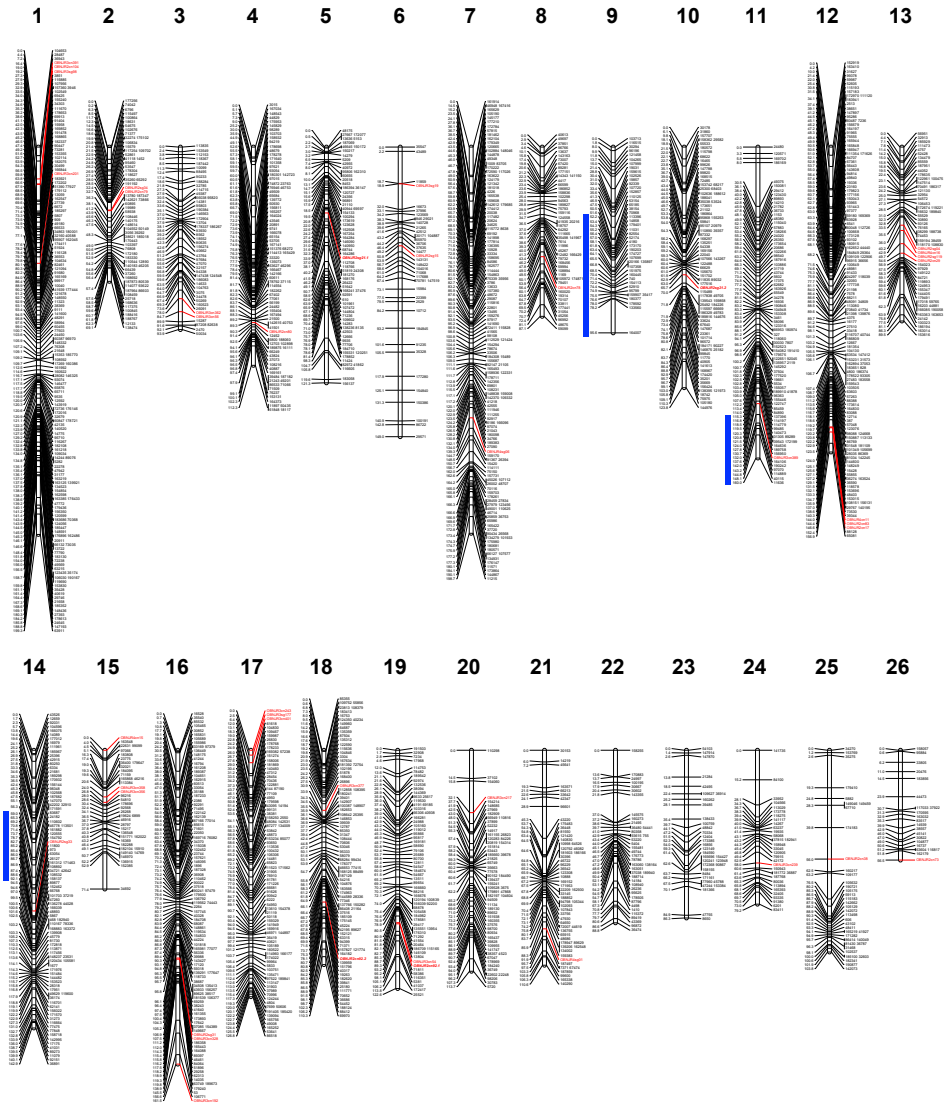


Fig 2. Sweet basil linkage map constructed for MRI x SB22 F_2 intercross family. The map includes 1,847 SNP (black font) and 42 EST-SSR markers (red font) across 26 LGs for a total length of 3030.9 cM. Two pairs of multi-locus EST-SSR (bold/italic/red font) markers represent putative homeologous pairs of loci (LGs 5,10 and 18,19). Blue lines represent 1.5 LOD score confidence intervals located to the left of linkage group locations associated with downy mildew resistance.

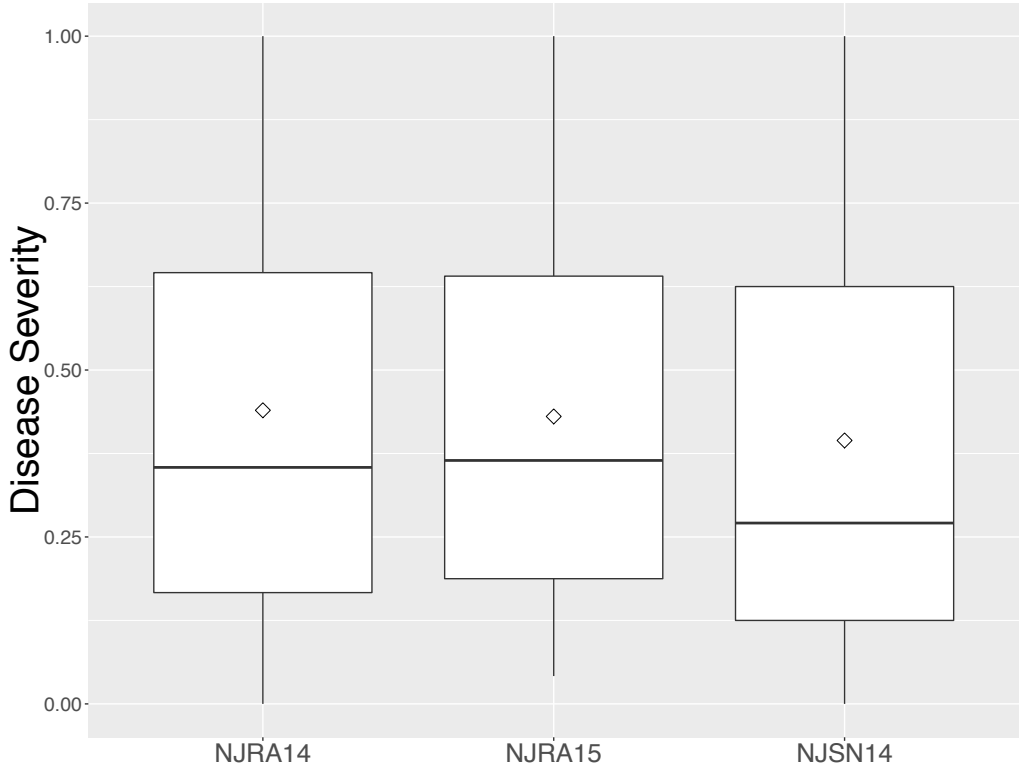


Fig 3. Frequency distribution of disease severity in the MRI x SB22 F_2 mapping population across three environments. Codes for each environment are shown on the x-axis and correspond to data recorded in 2014 in southern New Jersey (NJRA14), 2015 in southern New Jersey (NJRA15) and northern New Jersey in 2014 (NJSN14). Disease severity measured on a scale in which 0 = lowest possible severity score and 1 = highest possible severity score.

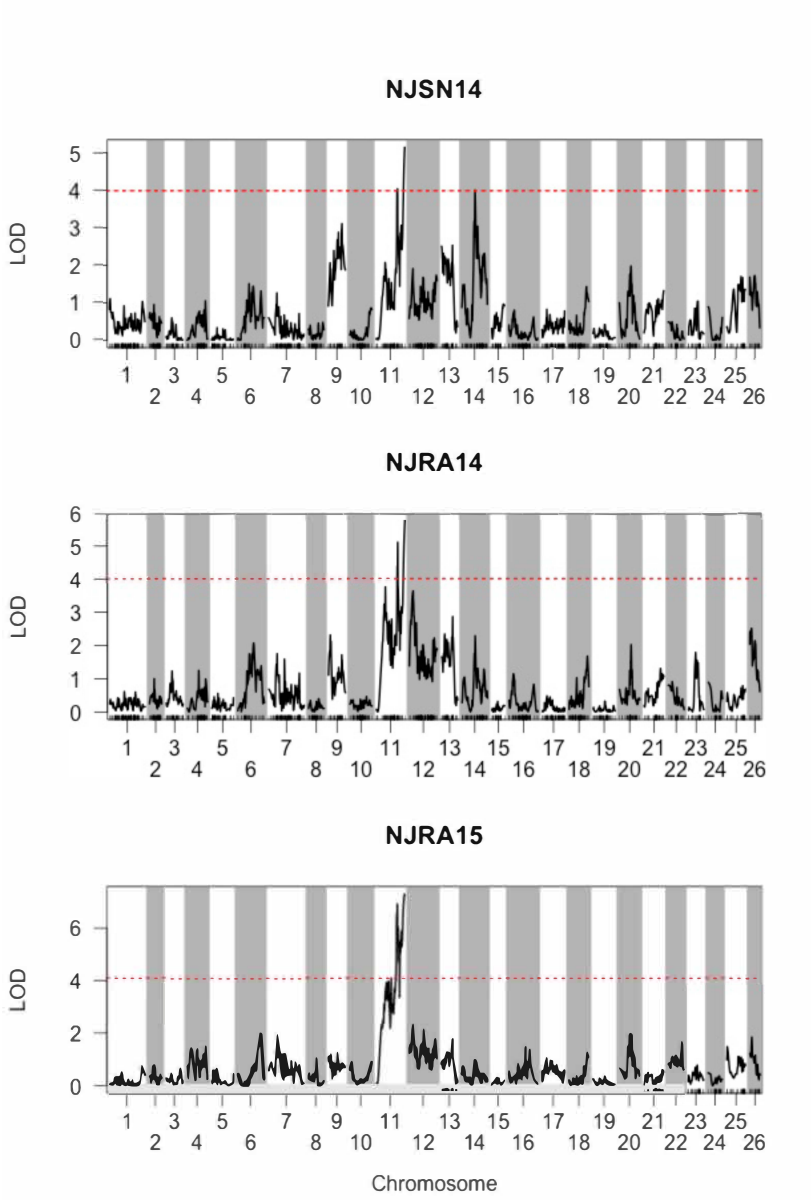


Fig 4. Detection of major downy mildew resistance QTL *dm11.1* across three environments. LOD scores for genome-wide scan using square-root transformed phenotype data from three environments: NJSN14 (northern New Jersey; 2014), NJRA14 (southern New Jersey; 2014) and NJRA15 (southern New Jersey; 2015). Significant LOD thresholds ($\alpha = 0.05$) were calculated by permutation tests with 1,000 iterations and are shown with red, dashed horizontal lines.

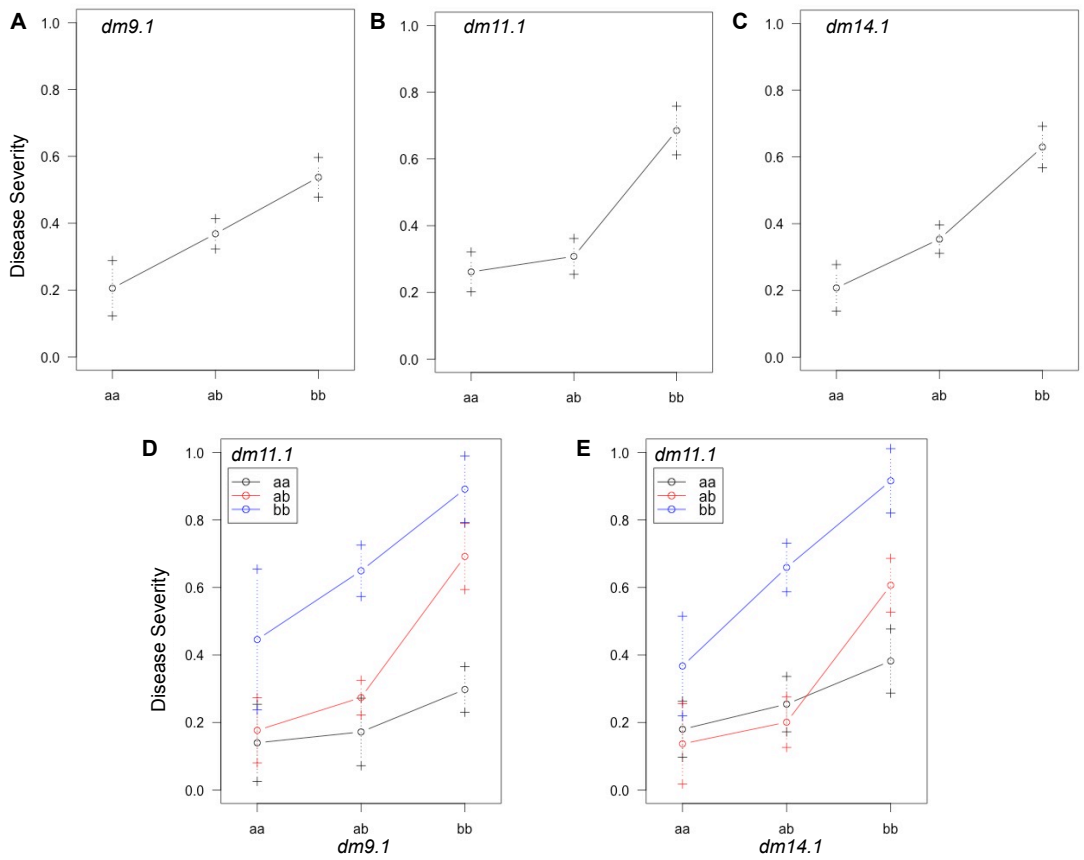
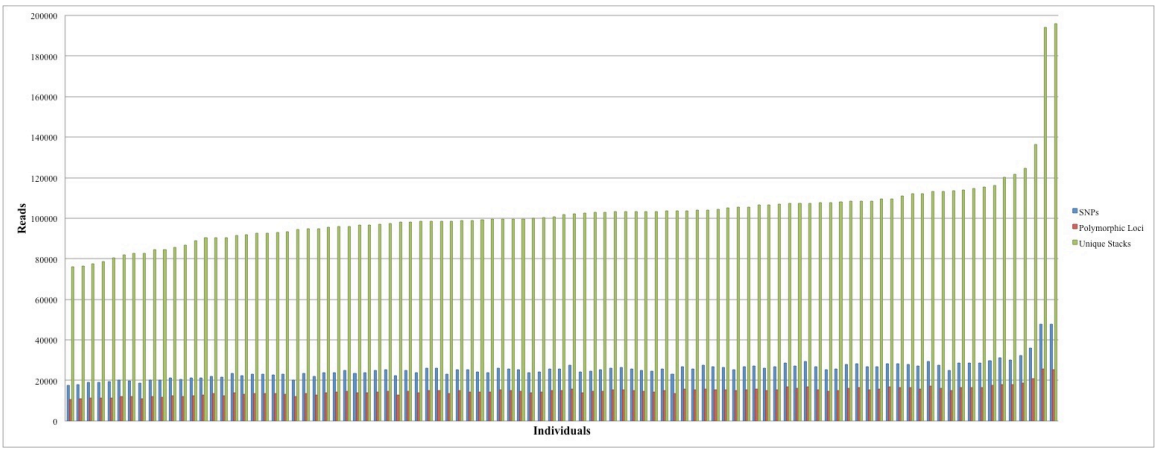


Fig 5. Effect and interaction plots for three QTL detected in environment NJSN14. MRI x SB22 F_2 genotype means (circles) ± 1 SE (error bars) for (A) minor QTL *dm9.*1, (B) major QTL *dm11.1* and (C) minor *dm14.1*. Two-QTL genotype by genotype means ± 1 SE for (D) *dm11.1* by *dm9.1* and (E) *dm11.1* by *dm14.1*. Allele 'a' is inherited from downy mildew resistant grandparent MRI and allele 'b' is inherited from susceptible grandparent SB22. Error bars represent ±1 SE.

Supplementary Table 1. Primer sequences for the EST-SSRs included in the MRI x SB22 F_2 linkage map.

Marker ID	NCBI Accession/ Contg ID	Forward Primer Sequence	Reverse Primer Sequence
OBNJR3cn391	Contig3280	CCCACCTCATCTTCTCATGG	CAGCTTGAAGTAGCCCTTGG
OBNJR2cn104	Contig3401	ATGGTGGGTTTAGCCATGAA	AAACAGCAGACAATTCGACAAA
OBNJR3sg98	DY337354	ACCAAATCCAAGACCCTCCC	TTGTAGAAGAGGCTCGTCGG
OBNJR3cn201	Contig1783	GCAGCAGCATTCAGGTACAA	GGGAGATTATTCACGAGGCA
OBNJR2sg34	DY331634	CCCAGGATTATTCCCTCATT	GAACATGGGAGGGATGAAGA
OBNJR2cn79	Contig2573	GGCGATGCTGGAGAACATT	GGAAAGTAGATCCGAGAGGGA
OBNJR3cn362	Contig2969	GAAGAGATGGCTGGTCTTGG	AGACAGAGAGAGGGCAGCAG
OBNJR3cn56	Contig582	GAAACAACATCCCTCATGCC	TTGAGATTGGGTTGGAGGAG
OBNJR2cn80	Contig2575	ATTTCAGCGCTCACATGACA	AGGAGCTGGATGGAAAGTCA
OBNJR2sg21	DY339566	TTTGCTCTGCTGGAGGGTAT	CAACAGGCATCGAAGTAGCA
OBNJR3sg19	DY343509	AAGCCGCCCTATAAACCAAT	GGCCGTTACAAAGAGCTGAG
OBNJR2sg15	DY340778	CAACTGCTAGTCGTGGGACA	CGACTCATGACCAGTAAACCTG
OBNJR4sg06	DY338242	CAAAGAGCCAATTAGTTTCCC	AGGCGACGGATTCATAGTTG
OBNJR2cn78	Contig2475	GGACAGAATTGCTACGAGGC	ATTGTGCTGCTGAACCCTTT
OBNJR3cn389	Contig3254	TGTCACCAAGAAGCATGGAG	CCATGACAAGTCGGGTTCTT
OBNJR4cn11	Contig1679	CTGTCATCGCCACAAGCTAA	TTGTGGCGCTTGAGAAGTTA
OBNJR2cn83	Contig2631	CTTCCGCAATCAGAAGAAGC	TGAATTTGTAGCGCACTTCG
OBNJR2cn17	Contig606	CTAGAGCTAGCGCAGGATGC	GATCGTCCGGTATTGCAGA
OBNJR2sg04	DY343638	ACGATATGAGACATGGGCCT	CGCAGGTACAAGCTTCTCAA
OBNJR4cn16	Contig2294	TTCACTCTGCCAGGCCTAAT	CTGTTTGAGCTGTGACGGAA
OBNJR3sg119	DY333250	ACACAGTAGATGCCGGTGGT	AAATGCTGGGCAAGAGTTTG
OBNJR2cn29	Contig1138	TGGACATCAAATTGGCTTCA	TGGAAGGACTCGTCATCTCTC
OBNJR2sg33	DY333933	GCCTCTCCCTCCTCCATAAC	AGGCGACGAGCATGAAGTAG
OBNJR4cn15	Contig2242	CAGCATCTCCGAACTGTGAA	AAACGATCATCTCCTCCACG

OBNJR2sg31	DY336298	CTTGAATTCGCGCAGTATGA	AAACAGCGGATTCACCACTC
OBNJR3cn328	Contig2750	CGTACAGCAGCAGTAGCAGC	GCTGCATTTGTGAACTGCTC
OBNJR3cn192	Contig1724	TCAGTGTTGAAGCTCGGAGA	TTCGATAGGAGGCTGAAGGA
OBNJR3cn243	Contig2153	ACTCTTGCTGTTTGTGGGCT	TCTTCAAAGTGGGCCAAGTC
OBNJR3sg177	DY322989	GGAATGTTACAAATGGCGCT	GTTGATGTGGATGTGGCTTG
OBNJR3cn401	Contig3437	ACCTGTAAACCAGCACCACC	TGACATGGGAGGAGGAACTC
OBNJR3cn377	Contig3126	CCCAAACAGGAAACAATAATTCA	CAACTCTTGAGCAGCGTTTG
OBNJR3cn54	Contig573	TTTCCCAACCCACTACCACA	CGGAAATGGGAATTAATTTGA
OBNJR2cn92	Contig3041	TGACATCAGCTCCAGAATGC	ACCCATATTTCGCCTTCTCA
OBNJR3cn217	Contig1936	ACTCCTTATGCTGGGACCCT	TCGTGCAGGAATGTGAAATC
OBNJR4sg01	DY343743	CAAACTTCAACCTCAACATTCAA	GAGGAGGAGGAGGAAGAGGA
OBNJR3cn239	Contig2134	CAAGGCAGCACAACATTCAG	AATGGCGTCTACCTTTGTGG
OBNJR2cn38	Contig1352	TCACGGTCAGCTCTCTCTCTC	CACACCGCTGAGTTTGAGAA
OBNJR2cn73	Contig2250	TAAGCCCTTTGGTCATCCAC	CAAGGACAATTCCTATTTAGTTCCA



Supplementary Figure 1. Stacks denovomap.pl output. Distribution of Stacks, SNPs and polymorphic loci identified in the MRIxSB22 F_2 population.

Dissertation Summary

The overarching goal of this dissertation was to provide a robust platform for the development of genetic resistance to downy mildew (*Peronospora belbahrii*) in sweet basil (*Ocimum basilicum*). Given the striking lack of available genomic information for sweet basil prior to the studies presented, this research also serves as a blueprint for mediating a destructive biotic stress in a non-model plant species. During the course of this research, distribution of the basil downy mildew pathogen (*Peronospora belbahrii*) extended throughout the world in the absence of genetic resistance. Results of three major studies provide needed answers to crucial questions regarding 1) *Ocimum* spp. genetic relationships, 2) heritability of *O. basilicum* downy mildew resistance and 3) location of resistance genes in the sweet basil genome.

A core set of 20 EST-SSR markers developed in this dissertation provides an important resource for breeding and phylogenetic research. Prior to this study, microsatellite (SSR) DNA markers had not been used to characterize relations among basil accessions, instead relying upon less reproducible RAPD and AFLP markers. A family-wide study had employed plastid markers, but included a very limited number of *Ocimum* spp. accessions. Accessibility to genic EST-SSR markers transferrable across laboratory experiments and basil species provides an important tool for improved breeding and phylogenetic classifications. EST-SSR genotyping of the *Ocimum* genus demonstrated distinct patterns of allelic distribution among populations determined by Bayesian model-based clustering. Occurrence of two and three alleles per locus within the inbred k1 *O. basilicum* and k2 *O. africanum* clusters, respectively, supported previous cytological evidence for allotetraploid and allohexaploid genomes. The third and most

genetically diverse k3 cluster included a number of exotic *Ocimum* species, the majority of which were highly resistant to downy mildew. Introgression of genetic resistance from k3 accessions into the elite germplasm of the *O. basilicum* k1 cluster is highly desirable but prevented by reproductive barriers. Genomic structure of accessions in the k3 cluster remains to be determined, but the occurrence of a majority single-allele SSRs indicates only one genome shares substantial homology with the tetraploid k1 cluster from which the EST-SSRs were derived. Does the k3 cluster include a diploid predecessor or are the accessions found in this cluster simply allopolyploid with sub-genome(s) of a more divergent ancestry? Further inference with respect to this matter will require additional investigation similar to that which has been performed in numerous allopolyploid species such as *Triticum spp*., *Gossypium spp*., *Brassica spp*. and *Fragaria spp*.

Results of phylogenetic analysis revealed relatively low k1 genetic diversity and in particular a single clade containing all commercially viable sweet basil varieties and breeding lines. This represents a potential genetic bottleneck and a need to broaden diversity among these accessions through introgression of more basal k1 accessions. Outcrossing was evidenced by mixed ancestry among the more basal, phenotypically diverse accessions of this cluster. Responsible breeding would work to introduce greater mixed ancestry among commercial sweet basil accessions. Stratification of the *Ocimum* genus provides the basis for additional studies such as phenotype-genotype correlation by genome wide association mapping. Care will need to be observed, however, to avoid false positive associations that could result from cryptic structure identified in the *O. basilicum* k1 cluster.

Identification of the novel, downy mildew resistant k1 accession, MRI, provided a feasible system for the development of a full-sibling family from a cross with downy mildew susceptible genotype SB22. Segregation of F_2 and backcross populations across two locations and two years provided clear evidence of dominant, major gene control of MRI-conferred resistance. This study provided the basis for a downy mildew resistance breeding strategy that has resulted in multiple inbred lines being nearing commercialization. Furthermore, phenotypic data provided preliminary evidence for (i) disomic inheritance and (ii) qualitative resistance. The former conclusion indicated *O. basilicum* adheres to the Mendelian laws of segregation and the latter that downy mildew resistance was only controlled by one or a few genes.

Predictable allelic segregation and major gene control of downy mildew resistance suggested genetic/QTL mapping of the SB22 x MRI F_2 mapping population to be feasible. A ddRADseq approach provided more than 1,800 SNPs with goodness-of-fit to the expected 1:2:1 genotype segregation model, facilitating the construction of 26 linkage groups with evenly distributed marker coverage. The mapping population was further genotyped with 48 EST-SSRs that demonstrated low levels of segregation distortion and two to four alleles per locus providing additional evidence for the *O. basilicum* k1 allotetraploid hypothesis previously proposed. EST-SSRs mapped to 24 of 26 linkage groups providing PCR-based anchor markers to allow for potential consensus map construction from future mapping populations. Ultimately, the validity of the map was confirmed through detection of a major QTL, *dm11.1*, across three different environments. Examination of gene action revealed a single MRI *dm11.1* allele conferred a resistance response. Multiple QTL mapping detected evidence of additional (minor)

loci having explained small, but significant portions of phenotypic variance. Results of QTL analysis were largely congruent with inferences made from statistical analysis of phenotypic data, but provided a more detailed measure of the loci involved and their relative effects. Furthermore, additional experiments such as fine mapping can be employed to convert the *dm11.1* QTL to a diagnostic marker for genomic assisted breeding.

Given the relatively short interval of time in which results were ascertained from these studies, this dissertation demonstrates the accessibility of breeding in a non-model plant system. It is important to note that the status of genotyping technology evolved rapidly during this same time. Strategies employed illustrate the importance of patience and careful survey of best practices or approaches to addressing a given hypothesis. For instance, awareness of the rapidly changing 'genomic landscape' through consistent monitoring of emerging technologies allowed for the identification of RADseq-facilitated genotyping. Prior to the development of this methodology the preferred strategy for a non-model system was exclusive SSR-based mapping that required exponentially greater cost, labor and time for a fraction of the markers (i.e. genome coverage). Computational requirements across all studies presented did not exceed that of a currently available CPU of reasonable processing power (~32 Gbp; ~2.2 GHz processor). In an era where big data has become the bottleneck in computational biology, incredible gains can still be achieved without the assistance of a server. Removal of the formerly prohibitive resource requirements for genotyping increases the accessibility of molecular plant breeding methods to non-model and specialty crop species.

A foundation is now established for implementation of modern breeding strategies in sweet basil. This will accelerate gain from selection with regard to resistance against the current and future downy mildew outbreaks as well as other stresses, biotic or otherwise. This dissertation research provides an experimental approach to elucidating the genetics of non-model species for applied plant breeding.