365313x, ja, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pj.16021 by Cornell University, Wiley Online Library on [2/1/1/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Pan-genome and multi-parental framework for high-resolution trait dissection in melon (*Cucumis melo*)

Elad Oren^{1,†}, Asaf Dafna^{1,2,†}, Galil Tzuri¹, Ilan Halperin¹, Tal Isaacson¹, Meital Elkabetz¹, Ayala Meir¹, Uzi Saar¹, Shachar Ohali¹, Thuy La³, Cinta Romay³, Yaakov Tadmor¹, Arthur A Schaffer⁴, Edward S Buckler^{3,5}, Roni Cohen¹, Joseph Burger¹ and Amit Gur¹*

¹Cucurbits Section, Department of Vegetable Sciences, Agricultural Research Organization, Newe Ya'ar Research Center, P.O. Box 1021, Ramat Yishay 3009500, Israel. ²The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, Faculty of Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel. ³ Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA. ⁴ Department of Vegetable Sciences Institute of Plant Sciences, Agricultural Research Organization, The Volcani Center, P.O. Box 15159, Rishon LeZiyyon 7507101, Israel⁵United States Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA. *Correspondence should be addressed to Amit Gur (e-mail: amitgur@volcani.agri.gov.il) †equal contribution

RUNNING TITLE: Melon pan-genome and trait dissection framework

KEYWORDS: pan-genome, *Cucumis melo*, structural variation, SNP, genetic mapping, halfdiallel, crop breeding, fruit-quality, disease resistance.

CORRESPONDING AUTHOR: Amit Gur, Newe Ya'ar Research Center, P.O. Box 1021, Ramat Yishay 3009500, Israel. <u>amitgur@volcani.agri.gov.il</u>.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/tpj.16021

SUMMARY

Linking genotype with phenotype is a fundamental goal in biology and requires robust data for both. Recent advances in plant-genome sequencing have expedited comparisons among multiple-related individuals. The abundance of structural genomic within-species variation that has been discovered indicates that a single reference genome cannot represent the complete sequence diversity of a species, leading to the expansion of the pan-genome concept. For highresolution forward genetics, this unprecedented access to genomic variation should be paralleled and integrated with phenotypic characterization of genetic diversity. We developed a multi-parental framework for trait dissection in melon (Cucumis melo), leveraging a novel pan-genome constructed for this highly variable cucurbit crop. A core subset of 25 diverse founders (MelonCore25), consisting of 24 accessions from the two widely cultivated subspecies of C. melo, encompassing 12 horticultural groups, and 1 feral accession was sequenced using a combination of short- and long-read technologies, and their genomes were assembled de novo. The construction of this melon pan-genome exposed substantial variation in genome size and structure, including detection of ~300,000 structural variants and ~9 million SNPs. A halfdiallel derived set of 300 F₂ populations, representing all possible MelonCore25 parental combinations, was constructed as a framework for trait dissection through integration with the pan-genome. We demonstrate the potential of this unified framework for genetic analysis of various melon traits, including rind color intensity and pattern, fruit sugar content, and resistance to fungal diseases. We anticipate that utilization of this integrated resource will enhance genetic dissection of important traits and accelerate melon breeding.

INTRODUCTION

Accepted Article

The ability to investigate gene functions in organisms has been revolutionized in recent years by the extensive implementation of genome editing in genetic research, through CRISPR-Cas9 technology (Doudna and Charpentier, 2014). Nevertheless, genetic dissection of simple or complex traits in crop plants is still fundamentally dependent on forward-genetics approaches, including the characterization of natural variation followed by genetic mapping of major Mendelian genes and quantitative trait loci (QTL). Genetic dissection of traits and QTL mapping have been accelerated by the rapid evolution of sequencing and genotyping technologies during the past two decades (Purugganan and Jackson, 2021). Parallelsequencing-based genotyping approaches supported by multiplexing protocols (Elshire *et al.*, 2011; Baird *et al.*, 2008) enable high-throughput and cost-effective dense genome-wide genotyping of populations. This technology-driven advancement facilitates efficient genotypic screening of large populations, and is shifting the bottleneck in mapping studies from genotyping to the phenotypic characterization of genetic diversity. Therefore, the development and availability of effective diverse germplasm and segregating populations for genetic mapping within species is an important factor.

Access to plant whole-genome assemblies in the early 2000s opened new avenues in crop genetics and further straightened the path from traits to QTLs and genes (Morrell *et al.*, 2011). The ability to use reference genomes for physical mapping and annotation of QTL intervals has been instrumental for moving from QTLs to causative genes. An important hidden layer in the genetic variation was only recently exposed through the assembly and comparisons of multiple within-species reference-genomes. The discovery of extensive intraspecific structural variation (SV) led to the realization that a single reference genome does not represent the diversity within

a species, and led to the expansion of the pan-genome concept (Della Coletta *et al.*, 2021; Bayer *et al.*, 2020).

Melons (Cucumis melo L., Cucurbitaceae) are among the most widely consumed fleshy fruits for fresh consumption worldwide (http://faostat3.fao.org/). Melons have been bred and are cultivated in nearly all of the warmer regions of the world, leading to the evolution of extensive diversity in phenotypic traits, especially in fruit traits such as size, shape, exocarp (rind) and mesocarp (flesh) color, sugar content, acidity, texture and aroma (Burger et al., 2006). This wide diversity is a source for further breeding and ongoing genetic research aimed at mapping QTLs and identifying genes affecting key horticultural and consumer-preference traits. Numerous genetic studies using diverse collections or segregating populations have focused on quality traits in melon, including fruit size and shape (Martínez-Martínez et al., 2022; Monforte et al., 2014; Oren et al., 2020), flesh color (Cuevas et al., 2009; Tzuri et al., 2015), rind color (Feder et al., 2015; Oren et al., 2019), netting and "sutures" (Oren et al., 2020; Zhao et al., 2019; Diaz et al., 2011), sweetness and aroma (Harel-Beja et al., 2010; Argyris et al., 2017), acidity (Cohen et al., 2014) and ripening behavior (Oren et al., 2022; Pereira et al., 2020; Ríos et al., 2017). Multiple analyses have also been performed for identification and genetic characterization of resistance to pathogens (Cohen et al., 2022; Burger et al., 2003; Joobeur et al., 2004).

Since 2012, mapping studies in melon have benefited from the availability of a reference genome (Garcia-Mas *et al.*, 2012) enabling physical mapping and annotations of QTL intervals. Resequencing of hundreds of melon accessions (S., Liu *et al.*, 2020; Zhao *et al.*, 2019) in recent years has further advanced the genomic analysis of candidate and causative genes. Five additional genome assemblies were published for *Cucumis melo* during the last 4 years. Three of these were of susp. *melo*: 'Payzawat' from the Inodorus Group (Zhang *et al.*, 2019), 'Harukei-3' from the Reticulatus Group (Yano *et al.*, 2020), and 'Charmono' from the Cantalupensis Group (Pichot *et al.*, 2022). The other two were of susp. *agrestis*: HS (Yang *et al.*, 2020) and IVF77 (Ling *et al.*, 2021). These assemblies provide additional references and important information regarding variation in *C. melo* genome structure, but a wider pangenomic view is still lacking for this species.

In the current study, we describe an integrated multi-parental framework that we developed for genetic dissection of traits in melon. This framework is based on a representative core subset of 25 diverse melon inbred accessions that was used to construct a wide set of 300 segregating bi-parental populations derived from a 25×25 half-diallele. A novel pan-genome that we built from the sequencing and *de novo* assemblies of the 25 genomes is integrated with the multi-parental germplasm resource to create a unified trait dissection framework. We provide here examples for the effectiveness and potential of this resource for genetic analysis of various phenotypic traits in melon.

RESULTS AND DISCUSSION

Selection of 25 founders from the GWAS180 diversity panel and development of the HDA25 and HDA25F₂ populations

Our primary resource in the current project is the diverse melon collection maintained at Newe Ya'ar Research Center. This collection contains many accessions drawn from the two widely cultivated subspecies of *Cucumis melo* (ssp. *melo* and ssp. *agrestis*) and includes representatives of 12 horticultural groups. This collection has been characterized for a wide range of phenotypic traits, and genotyped with ~24,000 genome-wide GBS-based SNP markers

(Gur *et al.* 2017). Integration of the combined subspecific/horticultural-group classification with the phenotypic and genotypic data facilitated the informed selection of a balanced core subset composed of 25 founders that represent the overall diversity of *C. melo* (hereafter *MelonCore25;* Figure 1a and b, Table 1; Gur *et al.* 2017). As such, the core set includes representatives of the two cultivated sub-species and the different horticultural groups in melon as well as the broad phenotypic spectrum available for key traits, as previously described (Gur *et al.*, 2017). The core set includes 17 accessions of ssp. *melo* (according to cultivar-group: 5 Inodorus, 4 Cantalupensis, 2 Reticulatus, 2 Khandalak, 1 Duda'im, 1 Flexuosus, 1 Ameri, and 1 thought to be Adzhur), 7 accessions of ssp. *agrestis* (according to cultivar-group: 2 Chinensis, 1 Conomon, 3 Makuwa, and 1 Momordica), and 1 feral accession, ssp. *collosus*, collected in central Israel.

To fully exploit the genetic diversity of this multi-parental resource, we developed a halfdiallel population by intercrossing the 25 founders in all possible combinations (*HDA25*, Dafna et al. 2021). This crossing scheme was further advanced by self-pollinations of all the F1s to produce 300 distinct F_2 populations, now available for genetic dissection of melon traits (hereafter *HDA25F₂s*, **Figure 1b**).

Sequencing, *de novo* assembly of 25 chromosome-scale genomes and characterization of structural genomic variation

Accepted Article

The 25 genomes comprising MelonCore25 were sequenced by Oxford Nanopore Technology (ONT) to an average depth of $70 \times (53-134 \times)$, generating a total sequence of 642.4 Gb with a median read length of 10Kb (4.5–24Kb) and a median read quality of 10.0 (9.2–10.6, Sup. Table 1). ONT reads were assembled into contigs and then polished using Illumina short reads from the previous whole-genome resequencing of these accessions (average depth $30\times$, (Dafna et al., 2021)). The mean N50 of contig size across the 25 genomes was 8.4 Mb and the average number of contigs per genome was 349.5 (median 273, Table 1). The contigs were scaffolded into pseudomolecules through a reference-guided approach using 'Harukei-3' (Yano et al., 2020) as the reference. Assembly statistics across the *MelonCore25* exposed variation in genome size ranging from 358 Mb (ESL) to 374 Mb (BEL), with a mean of 365 Mb (Table 1). These assembly statistics are comparable to the two published high-quality melon reference genomes, DHL92 v4.0 (Castanera et al., 2020) and 'Harukei-3' (Yano et al., 2020). Additional quality benchmarks for our assemblies are the high gene-space integrity and completeness, based on BUSCO scores that ranged from 97.5% to 98.1% complete genes, and the completeness of transposable elements based on the LTR Assembly Index (LAI) with an average of 12.9 (Table 1). The 25 genomes were annotated for gene models and using stringent parameters (described in Experimental Procedures section), we found variation ranging from 36,175 to 37,259 genes per genome, that was not correlated with genome size (Table 1). Interestingly, genome size variation was correlated with the sub-specific division such that the genomes of ssp. agrestis accessions were significantly smaller (mean 361 Mb) than those of the ssp. melo accessions (mean 366.3 Mb, Figure 2a, b). Alignment of the 25 genomes and analysis of structural variation (SV) revealed extensive polymorphism in all types of SVs (InDels, inversions and translocations). Large intra-chromosomal inversions spanning across 1.6 Mb and 3.2 Mb that differentiated between most agrestis and melo lines were found in chromosomes 1 and 11, respectively (Figure 2a). The independent discovery of these inversions across multiple accessions is further evidence for the reliability of our assemblies. We identified 108,000 large SVs (InDels and translocations) ranging from 2 Kb to 1,300 Kb (Figure 2c, Sup. Table 2). Sixty-three percent of these, 68,500 SVs, were rare and 39,500 SVs were present in two or more accessions (Figure 2d). We then focused on characterization of short InDels that are useful as simple PCR-based genetic markers in mapping studies. We identified 190,543 short InDel alleles in a size range of 10–2,000 bp (Figure 2e, Sup. Table 3). A significant proportion of these InDels represent multi-allelic polymorphisms and converged into 115,802 sites. Allele frequency analysis across the 86,669 bi-allelic sites showed a similar distribution pattern as in the large SVs, such that almost half of these InDels were rare (Figure 2f). To confirm that *MelonCore25* comprehensively represents the genomic diversity in melon, a pan-genome analysis for the 86,669 bi-allelic InDels was performed by plotting the number of cumulative InDels against the number of genomes. The results showed that 22 accessions were sufficient to capture 98% of the SV (Figure 2g). The short InDel profiles of the parents were then projected on the 300 HDA25 crosses to obtain a detailed description of the polymorphisms in each cross (Sup. Table 3). The number of polymorphic sites per cross range from 5,082 to 45,537 and, as expected, was strongly correlated with parental genetic distance previously calculated based on 24,000 GBS SNPs (Figure 2h). Short-InDels coverage across the 12 melon chromosomes was relatively uniform in the MelonCore25 pan-genome and specific crosses, regardless of the mating distance (Figure 2i). Even in the cross between the two closely related agrestis accessions, OHG and SAS, we identified more than 400 short InDels per chromosome. A random set of 22 short InDels from five of the chromosomes were validated by PCR (Sup. Figure 2, Sup. Table 4).

Genome sequences are available to date for 18 species of *Cucurbitaceae*. These cucurbit genomes exhibit much variation in size and organization (Ma *et al.*, 2022). Information on intra-specific variation in genome size within this family is, however, still relatively limited. Genomes of 11 *Cucumis sativus* L. (cucumber) accessions were recently sequenced and compared, exposing a difference of 26 Mb (11%) between the largest and smallest genomes, as well as 56,214 SVs (Li *et al.*, 2022). We observed size variation of 16 Mb (4.5%) between the largest and smallest genomes across *MelonCore25*. The difference in genome size was significant between the *melo* and *agrestis* sub-species (Figure 2b), and imply on the possible evolutionary significance of within-species genome size variation (Biémont, 2008). The variation in genome size and the consistently differential, large chromosome inversions (Figure 2a) imply that these two subspecies have evolved separately under cultivation, with little or no interaction. Parallel evolution under domestication for fruit-quality traits has also been described for the two cultivated subspecies of another cucurbit, *Cucurbita pepo* L. (squash/pumpkin) (Paris *et al.*, 2012).

Workflow for utilization of the multi-parental framework

Accepted Article

Our integrated framework relies on the connection of the structured multi-parental set of populations ($HDA25F_{2}s$) with phenotypic data, whole-genome assemblies and the derived variome of the 25 parents (**Sup. Tables 2 and 3**). The proposed utilization workflow is described in **Sup. Figure 1**. Briefly, a trait of interest should first be scanned across *MelonCore25* by either using the database for the trait, or by direct phenotypic analysis. Following reliable phenotyping, we found that projecting the phenotypic data on the 2D genetic PCA plot of *MelonCore25* (coordinates available in **Sup. Table 1**) is a useful way to integrate genotype and phenotype for informed selection of parents differing in the trait of interest and reflecting varying parental genetic distances. Following the selection of parents, an optimal F_2 population can be obtained from the $HDA25F_{2}s$ set for characterization of the mode of inheritance and genetic mapping of the trait. Depending on the genetic architecture and complexity of the trait, the relevant mapping strategy can be applied and, accordingly, the population size and generation can be adjusted. Parental genome assemblies of the selected

population are available as reference for bulk-sequencing analyses (BSA-Seq, QTL-Seq) and for marker development for further fine-mapping. Detailed annotation of simple or quantitative trait locus (QTL) intervals can also benefit from referring to the specific parental genomes. After the initial whole-genome mapping, the $HDA25F_{2}s$ set can be a valuable resource for selection of additional relevant crosses for QTL validation or allelism testing. Genomes of the 25 founders can then be used for mining allelic variation at candidate genes.

Examples for implementation

Accepted Article

Genetics of pigment intensity and mottled rind

Mottled rind is a common characteristic in melon and is expressed in different patterns across our GWAS180 collection (Figure 3a). As with other traits, MelonCore25 captures this diversity with parallel proportions of mottled and non-mottled rind, as in the wider GWAS180 collection (Figure 3b). To present a comprehensive description of its genetic architecture, the rindmottling trait was projected on the genetic PCA plot and several mottled and non-mottled accessions were selected for focused genetic analyses (Figure 3c). First, mottled rind was mapped, using the cross between 'Dulce' (DUL, ssp. melo, Reticulatus Group), which has a uniformly dark rind, and the mottled-rind PI414723 (PI414, ssp. agrestis, Momordica Group), using a previously described RILs population (Figure 3d; (Galpaz et al., 2018)). Wholegenome linkage analysis resulted in the mapping of a single major trait locus to ~400Kb interval on chromosome 2 (Figure 3e). The association of this locus with mottled rind was validated through independent GWA analysis of this trait on our large collection (Figure 3f, GWAS180, n=177). The integration of linkage and association analyses allowed us to narrow the mapping to a common genomic interval of less than 200Kb, which was further confirmed through substitution mapping using a detailed analysis of recombinants obtained from the PI414×DUL RILs (Figure 3g).

To expand the analysis of the mode of inheritance of the mottled-rind trait, rind pattern of young fruits was examined across the HDA20 population (subset of 20 founders from MelonCore25, and their 190 half-diallel F1s). All crosses with a mottled parent resulted in mottled F_1 s and non-mottled F_1 s were obtained only by the combination of two non-mottled parents (Figure 4a). The results from F_2 populations were also consistent with a single dominant gene for mottled rind. Subsequently, two additional segregating populations from crosses between mottled and non-mottled parents were examined. One was an F_2 population from the cross between the mottled ssp. *collosus* accession QME and the non-mottled, ssp. *melo*, Inodorus Group TAD (Figure 4b). The other was an F_2 population from the cross between two closely related ssp. agrestis accessions, the mottled PI161375 (Chinensis Group) and the uniformly light, non-mottled ESL (Makuwa Group) (Figure 4c). In both populations, rind mottling was phenotyped on young fruits (10-15 days post anthesis). A significant association was observed between the markers that we developed at the MT-2 locus and mottled-rind segregation (Figure 4b, c). Interestingly, segregation and association patterns in these populations were different when compared with the PI414×DUL RILs population (Figure 4e). In both crosses, groups homozygous to the non-mottled allele at the MT-2 marker were perfectly associated with non-mottled rind, but the heterozygotes and homozygotes for mottledrind allele included also significant proportion of non-mottled rind segregants (Figure 4 b, c, e). Such pattern suggests that epistasis with another locus is affecting the expression of this trait. Recently, (Shen et al., 2021) proposed that CmAPRR2 is epistatic over MT-2 in controlling

mottled rind phenotype in a cross between the mottled *agrestis* inbred ('Songwhan Charmi') and a non-mottled *ssp. melo* 'Mi Gua'. The PI161375×ESL population displayed clear segregation to light and dark rind color and was therefore also used to test the association between CmAPRR2 gene and rind pigment intensity in mature fruits. In a previous study, we proposed that the light rind of ESL is a result of a single-base insertion in exon9 of the CmAPRR2 gene, causing a frameshift that leads to a major modification in the predicted protein sequence (Oren *et al.*, 2019). PI161375 carries the normal ("dark") allele of this gene and, indeed, complete co-segregation was observed between the SNP in CmAPRR2 and rind color, supporting the proposed exon9 polymorphism as causative (**Figure 4d**). When combined, the genotypic data from the CmAPRR2 and the MT-2 markers in the ESL×PI161375 population confirmed that the segregation of the two genes can perfectly explain the mottled and non-mottled rind segregation (**Figure 4f**), with the light allele of CmAPRR2 epistatic over MT-2.

To further validate the assumed allelism among the various mottled-rind sources, three additional crosses were made between mottled lines. One was between DUD (ssp. *melo*, Duda'im Group) and PSR (ssp. *melo*, Inodorus Group) (Figure 4g). Another was between QME (ssp. *collosus*) and PSR (Figure 4h), and the third between QME and PI414 (Figure 4i). In all three crosses, 100% of the F_2 segregants (n= ~100) displayed mottled-rind fruits, providing further supportive evidence that the various mottled rind patterns across melon diversity are allelic and conferred by the *MT-2* gene.

Rind color and pigmentation pattern are primary components of melon fruit appearance. We previously reported CmAPRR2 gene as a key regulator of rind and flesh pigment intensity in melon (Oren *et al.*, 2019). That study benefitted from an integrative analysis of pigment variation across multi-parental germplasm. Genomes resequencing of MelonCore25 were used for a detailed analysis of haplotypic variation in the CmAPRR2 gene and allowed detection of multiple independent predicted causative mutations across the diversity, supported by comprehensive allelism tests using the HDA25 population. This high-resolution multi-parental analysis exposed the complex haplotypic variation in the CmAPRR2 gene and resolved the unexpected lack of signal in GWAS for this conserved central pigment accumulation regulator (Oren *et al.*, 2019). Co-segregation analysis between the light-rind phenotype and the ESL allele in the current study provides further support for the multi-allelic model and predicted causality of an exon9 point mutation in CmAPRR2 for the light rind of ESL (Figure 4d).

The mapping of the mottled rind trait was accomplished using a combination of segregating populations and a diverse collection (Figure 3 e, f). The trait has been mapped to a ~ 115 Kb interval with 14 annotated genes. These results coincide with previous mapping of the MT-2gene by parallel analyses of different crosses which supported the probable allelic nature of the mottled-rind trait from diverse sources. Lv et al., (2018) mapped the "spotted" rind trait in the Chinensis Group to 280 Kb interval on chromosome 2. Pereira et al., (2018) described the mapping of mottled rind to the same genomic region in a cross between the non-mottled rind 'Védrantais' (ssp. melo, Cantalupensis Group) and the mottled 'Piel de Sapo' (ssp. melo, Inodorus Group). Recently, Shen et al., (2021) mapped mottled rind to a 40.6 Kb within the same interval harboring 6 genes, using a ssp. agrestis Chinensis Group accession as the trait donor. Conclusive evidence for the causative gene and sequence variants is yet to be provided. The mapping interval reported here overlap with the 40.6 Kb interval of Shen et al. (2021) and the detailed genomic variation available now for MelonCore25 can assist in identifying the causative variation. We analyzed SNPs and SVs within the 6 candidate genes in the 40.6 Kb interval (Sup. Tables 5 and 6). This region possesses 62 exonic SNPs, 21 of which are nonsynonymous polymorphisms, 14 are untranslated region (UTR) polymorphisms and 27 are synonymous or splice-site polymorphisms. Interestingly, none of these polymorphisms display

a significant association with the mottled-rind phenotype across the *MelonCore25* set. The broad multi-parental analyses here and in previous studies support the notion that mottled-rind variation is determined by a single gene across melon diversity which could suggest a possible monophyletic origin of this trait. However, the lack of association of an exonic polymorphism across the 25 founders could also suggest that the variation is either regulated on other levels (such as expression or translation), or by multi-allelic variation caused by different independent mutations in the causative gene, as previously described for the *CmAPRR2* gene (Oren *et al.*, 2019). Resolution of this riddle could be achieved once the causative sequence variation is discovered and characterized.

Genetics of sugar accumulation

Article

Accepted A

Fruit sweetness is the primary parameter defining melon quality and consumer preference. Much research has been devoted to obtaining improved understandings of carbohydrate metabolism and the complex genetic architecture of sugars accumulation and fruit sweetness (Dai et al., 2011; Harel-Beja et al., 2010; Schaffer et al., 1987; Leida et al., 2015; Argyris et al., 2017). MelonCore25 has been characterized for total soluble solids (TSS, °Bx) content of mature fruits grown in replicated open-field experiments. Mean TSS of the 25 accessions varied from 4 to 15 (Figure 5a). This variation is similar to the range displayed across our broader *GWAS180* collection (Gur *et al.*, 2017) and reflects the wide spectrum of fruit sweetness within *Cucumis melo* and the relevance of studying the genetics and physiology of sugar transport and accumulation within this species. Projection of TSS values on the genetic PCA of MelonCore25 (Figure 5b) shows that fruit sweetness is not independent of population structure, as most of the ssp. *melo* accessions are sweet, with the exception of non-sweet lines from the Flexuosus and Adzhur Groups, but the ssp. agrestis accessions in our core set contain a more equal representation of high and low TSS. Based on this integrated view, efforts were focused on several crosses between high and low TSS accessions that represent different combinations of subspecies and cultivar-groups. The rationale was to characterize the genetic architecture of sugar accumulation in multiple crosses that represent melon diversity, which could facilitate a comprehensive view of the genetic architecture of fruit sweetness. For example, HDA232 is a cross between two closely related ssp. agrestis accessions differing in TSS: ESL (Makuwa Group, Bx=14.6) and PI161375 (Chinensis Group, Bx=7.0), and therefore such a cross potentially represents allelic variation in a small number of major sugar QTLs. On the other hand, HDA243 is an inter-subspecific combination between genetically distant accessions, the sweet ssp. agrestis SAS, (Makuwa Group, Bx=14.0) and the non-sweet ssp. melo DOYA (Flexuosus Group, Bx=3.8), and is therefore expected to reflect variation in a large number of sugar QTLs. Examples of three selected segregating populations that were advanced for mapping sugars QTLs are provided in Figures 5d-f. The TSS correlation between F3 and F4 family means from the cross between TAD (ssp. melo, Inodorus Group, Bx=15.2) and the nonsweet QME (ssp. collosus, feral, Bx=8.0) is presented in Figure 5d, with tail segregants from 208 F3:F4 families highlighted for QTL-Seq analysis. Figure 5e shows the TSS correlation between F4 and F5 family means from the cross between TAD and a different non-sweet accession, PI164323 (subsp. melo, Adzhur Group, Bx=4.9). Tail segregants from this population were also selected for mapping the main sugar QTLs in this cross. Figure 5fpresents the variation in TSS and tail segregants selection in the F5 population derived from the intersubspecific cross between the sweet SAS and the non-sweet DOYA.

To expand our view on the inheritance of sugar accumulation using the multi-parental framework, we analyzed TSS in the HDA20 sub-population composed of 20 parents from

MelonCore25 alongside their 190 half-diallel F1s. In order to gain broad insight on the mode of inheritance of fruit TSS, each of the 190 F1s was compared to its parents. Various modes of inheritance were observed across the broad *HDA20* set (**Sup. Figure 3**); however, the highly significant correlation (r = 0.81) between mid-parent and F1 TSS (Figure 5g) across the 190 diverse hybrid groups is indicative of a mostly additive inheritance of TSS through melon diversity. Taken together with the complex genetic architecture of fruit TSS, these results imply that the use of multiple bi-parental populations could be an effective path to genetically dissect this trait and develop marker-assisted selection protocols to breed sweet melons.

In addition to TSS measurements, *MelonCore25* was also analyzed for sugar composition. As previously shown (Burger and Schaffer, 2007; Schaffer *et al.*, 1987), the sugar content continuum in melon is explained by the variation in sucrose rather than in its hexoses building blocks—glucose and fructose (Figure 5c). This unidirectional variation, reflected through the bivariate view of the hexoses versus sucrose content across *MelonCore25* (Figure 5c) is another approach to express variation in sugar composition. Of the total sugars in melon fruit flesh, there is a highly variable proportion of sucrose, with the sweet-fruited accessions containing 85% sucrose and the non-sweet only 15%. This variation could prove to be an effective substrate for establishing genetic studies focused on sugars metabolism in melon fruit.

Although there is a clear distinction between sweet and non-sweet melon cultivar-groups, previous studies suggest that the genetics of sucrose metabolism and accumulation is complex and most likely controlled by multiple QTLs with small effects rather than variation in major genes (Diaz et al., 2011; Leida et al., 2015). The physiological and biochemical characterization of sugar metabolism performed thus far in melon has provided many candidate genes related to sugar metabolic pathways and transport (Dai et al., 2011), but causative genes explaining the phenotypic variation are yet to be discovered through effective genetic analyses. Argyris et al. (2017) analyzed multiple segregating populations from two crosses between sweet and nonsweet parents, for mapping sugar-accumulation QTLs. This approach is expanded here by targeting multiple crosses from the $HDA25F_{2}s$ set. We selected several crosses between highand low-sugar parents from several inter-subspecific and intra-subspecific combinations (Figure 5b). Using the genome assemblies of all the participating parents in these crosses, parallel QTL-Seq analyses can be performed using tail segregants (Figure 5 d-f) and efficient annotations of sugar accumulation QTL intervals can be anticipated in these populations. In addition to the complete representation of SNP and short InDels variation in these populations, a detailed comparative analysis of structural variation within QTL intervals or presence-absence variation (PAV) analysis of sugar metabolism and transport genes can be further analyzed and associated with phenotypic variation.

Genetics of resistance to soil-borne pathogens

Accepted Article

Variation in resistance to Fusarium wilt and haplotypic variation in Fom-2 gene across MelonCore25

Fusarium wilt is an important disease affecting melon production, caused by the soil-borne fungus *Fusarium oxysporum* f. sp. *melonis* (*F. o. melonis*). We characterized the response of young seedlings of the *MelonCore25* set to artificial inoculation with *F. o. melonis* races 1 and 2. For both pathogen races, a wide spectrum of responses from resistant to susceptible was observed and there was, as expected, a differential response to the two races (**Figures 6a** and **6b**). Projection of the disease-symptom indices on the genetic PCA displayed a division of response to race 1 by subspecies, such that the ssp. *agrestis* accessions were resistant and all

but one of the ssp. *melo* accessions were intermediate or highly susceptible. The differential response to the races was most apparent among the *agrestis* accessions, that were resistant to race 1 and highly susceptible to race 2. A major gene (Fom-2) conferring resistance to F. o. melonis races 0 and 1 has been mapped and positionally cloned (Joobeur et al., 2004). Genetic marker was developed within the Fom-2 gene for marker-assisted selection in melon breeding (Wang et al., 2011). We used our multi-parental platform to evaluate cases of limited predictive value of the Fom-2 marker observed across different breeding germplasm. The 25 genome assemblies allowed us to obtain a detailed alignment of the Fom-2 (Melo3C021831) gene sequence and detect multiple polymorphisms, including a large InDel between exon2 and exon3 (Figure 6c). In all, we found 45 polymorphic sites within exons of the *Fom-2* gene, 26 of which were non-synonymous polymorphisms resulting in different types of effects on the expected protein sequence (Sup. Table 7). Interestingly, this extensive sequence diversity is translated into only 3 main haplotypes across the *MelonCore25* set. Haplotype 1 is specific and common to the ssp. agrestis accessions in our core set and is similar to the DHL92 reference in this region (Figure 6c, d). Haplotype 2 is distributed across the different cultivar-groups within ssp. melo, and characterized by multiple SNPs along the gene that are part of a linkage disequilibrium (LD) block. Haplotype 3 is common to ssp. melo lines from different cultivargroups and contain an insertion of $\sim 1,100$ bp compared to the reference, and a few additional SNPs in the LD block (Figure 6c, d). While we found a significant association between these Fom-2 haplotypes and Fusarium race 1 resistance across MelonCore25, none of the polymorphisms independently explained the variation (Figure 6). The only exception to this haplotypic association is VEP (ssp. melo, Cantalupensis Group) (Figure 6c: Haplotype 2, resistant to race 1), suggesting that additional genes are involved. The cross between VEP and the related susceptible DUL (ssp. melo, Reticulatus Group, Haplotype 2) is therefore an informed selection of a potential population for mapping these other genetic loci (Figure 6a).

Genetic analysis of resistance to Fusarium oxysporum f. sp. radicis-cucumerinum (FORC)

Accepted Article

Fusarium root and stem rot caused by the fungus Fusarium oxysporum f. sp. radiciscucumerinum (FORC) is a disease of greenhouse-grown cucumbers and melons (Punja and Parker, 2000; Vakalounakis et al., 2005). Genetic variation in the response to artificial inoculation with FORC has been described (Cohen et al., 2015; Elkabetz et al., 2016). Inoculation of FORC across the MelonCore25 set revealed a wide spectrum ranging from susceptibility to resistance (Figure 7a). Projection of disease severity index (DSI) on the genetic PCA plot showed that the level of resistance to this pathogen is independent of population structure (Figure 7b). Multiple potential crosses between resistant and susceptible accessions are available in the $HDA25F_{2S}$ set, and we used the RILs population developed from the cross between TAD (ssp. melo, Inodorous Group, intermediate resistance) and DUL (ssp. melo, Reticulatus Group, susceptible) (Oren et al., 2019) to genetically dissect this trait. The RILs population displayed transgressive segregation in response to FORC inoculation, with lines more resistant than TAD or more susceptible than DUL (Figure 7c). This population was genotyped with 89,343 GBS SNPs defining 2853 recombination bins (Oren et al., 2019) and through whole-genome QTL mapping analysis we identified a major QTL on chromosome 7 (Figure 7d) that explains 22% of the FORC DSI variation ($P = 2 \times 10^8$, Figure 7e). We used the parental de novo assembled genomes and derived melon variome database to extract informative InDel markers for further backcrossing to validate and fine-map this QTL.

Variation in response to charcoal rot (Macrophomina phaseolina) across MelonCore25

We recently completed a multi-environment screen of *MelonCore25* for resistance to the soilborne fungal pathogen *Macrophomina phaseolina* which causes a disease often referred to as charcoal rot (Cohen *et al.*, 2022). The frequency distribution of cross-environment average disease severity index (DSI) of the 25 accessions is plotted in **Sup. Figure 4a.** The whole spectrum from resistant to highly susceptible was observed across *MelonCore25*. Projection of the DSI values on the genetic PCA (**Sup. Figure 4b**) highlight the correlation between intraspecific classification (ssp. *melo* and *agrestis*) and resistance level, such that most ssp. *agrestis* showed tolerance or moderate susceptibility, while most ssp. *melo* showed moderate or high susceptibility. Potential crosses are available between the tolerant and susceptible accessions in the $HDA25F_{2S}$ set, and as such are a good substrate for genetic dissection of this trait.

A unified framework for trait dissection in the pan-genomic era

As many horticulturally important traits are polygenic, the availability and utilization of multiparental resources is required for comprehensive genetic dissection of natural variation and detection of a repertoire of favorable alleles for crop breeding. Melon genetic and genomic research has been much enhanced since the release of the first melon genome 10 years ago (Garcia-Mas et al., 2012), followed by extensive re-sequencing of hundreds of melon accessions (Zhao et al., 2019; S., Liu et al., 2020; Dafna et al., 2021). Five additional highquality melon genome assemblies that can be used as references for this species were published over the past 4 years (Zhang *et al.*, 2019; Yang *et al.*, 2020; Yano *et al.*, 2020; Ling *et al.*, 2021; Pichot et al., 2022). However, the recent expansion of the pan-genome concept, based on the realization that multiple genomes capturing the diversity within species are required for complete representation of the genomic variation (Li et al., 2022; Della Coletta et al., 2021; Bayer et al., 2020; Ho et al., 2020), suggest that the SV layer is not fully exposed in melon. We, therefore, developed and describe here a multi-parental trait dissection framework, anchored to a novel pan-genome constructed based on high-quality assemblies of 25 genomes, representing melon diversity. Pan-genomes have been described for several crop plants, highlighting the commonality and magnitude of within-species structural variation (Li et al., 2022; Qin et al., 2021; Y., Liu et al., 2020; Jayakodi et al., 2020; Song et al., 2020; Gao et al., 2019). The unique aspect of the current project is the creation of a unified common framework that integrates genetic diversity, phenotypic variation, and genomic information. In addition to the traits described here, the MelonCore25 set was phenotyped to fruit morphology traits (size and shape), rind and flesh color and pigment content (Gur et al., 2017), rind netting (Oren et al., 2020), fruit firmness, ripening behavior (Oren et al., 2022), and over 80,000 metabolomic and elemental features extracted from rind and flesh (Moing et al., 2020). Alongside the 25 genome assemblies and segregating populations covering all possible parental combinations, the genetic basis of these traits (and others) can now be studied more effectively. An important additional layer that can supplement the current framework is a comprehensive expression profiling (transcriptomes) of the MelonCore25 accessions across different organs, tissues and developmental stages, which could provide an important layer for candidate-gene discovery and improve genomes annotations.

Conclusions

We describe here a new resource for genomic research in melon. The pan-genome constructed from 25 diverse accessions selected to represent the broad variation across *Cucumis melo* is

useful for comparative genomic analyses to address the evolution of the melon genome and the impact of breeding history. The integral connection of the pan-genome with a set of crosses and segregating populations ($HDA25F_2$, n = 300) is expanding the usability of this resource. For example, correlation between structural variation and parental crossability or fertility of F1s can be studied. Detailed analysis of relations between large genomic rearrangements and recombination frequency can be directly addressed using the SVs database and F_2 populations. We demonstrate the primary potential of this integrated framework: Genetic and genomic dissection of traits. The genomic information could also be useful for informed design of genome editing and trait introgression in breeding programs, taking into consideration the impact of SVs.

EXPERIMENTAL PROCEDURES

Plant materials and field trials

<u>MelonCore25</u>: This research is centered on the core set of 25 diverse melon accessions (**Table 1, Figure 1**) that were selected based on a comprehensive genotypic and phenotypic characterization of our broader *GWAS180* panel (Gur *et al.*, 2017). The accessions for this core set were selected based on the multiple genotypic and phenotypic criteria previously described (Dafna *et al.*, 2021; Gur *et al.*, 2017). Briefly, an initial tentative set (n = 40) was constructed to represent all the cultivar-groups of melon (Pitrat *et al.*, 2000; Burger *et al.*, 2009). Phenotypic profiles were then used as the second primary factor; the preliminary core set was projected on the distribution of the different traits to ensure a phenotypic spectrum that is well-captured in the core panel (as illustrated by Gur *et al.* 2017). Following required adjustments and narrowing of the set to n = 30, based on the first two steps, the final set was selected to meet the 25-accession target, taking into account maximum polymorphism information content (PIC) value and uniform distribution on genetic diversity plots. Genetic PCA coordinates of the 25 founders are available at **Sup. Table 1**.

<u>HDA25</u>: The creation of the diverse, 25-way half-diallel population, resulting in 300 F1 hybrids that represent all possible parental combinations, was previously described by Dafna *et al.* (2021).

<u>HDA20</u>: A subset of HDA25, HDA20 is composed of 20 representative lines of the 25 members of HDA25. The 20-way half-diallel population of HDA20 resulted in 190 F1s. This subset was described by Dafna *et al.* (2021).

<u>HDA25F₂s</u>: All 300 HDA25 F1s were grown in a greenhouse at Newe Ya'ar and subjected to self-pollinations to produce 300 F_2 populations.

<u>PI414723 × Dulce RILs population (*HDA019*, F7):</u> A bi-parental population of 99 RILs from a cross between the mottled-rind PI414723 (ssp. *agrestis*, Momordica Group) and the non-mottled-rind 'Dulce' (ssp. *melo* Reticulatus Group) was previously described (Harel-Beja *et al.*, 2010).

<u>Tam Dew × Dulce RILs population (HDA034, F_2)</u>: A bi-parental RILs population of 164 RILs from a cross between 'Tam Dew' (TAD; ssp. *melo* Inodorus Group) and 'Dulce' as described by (Oren *et al.*, 2020).

<u>Tam Dew × Qishu Meshullash (*HDA008*, F_6)</u>: Tam Dew was crossed with Qishu Meshullash (QME, ssp. *Colossus*), in a greenhouse at Newe Ya'ar and ~200 lines were advanced to the F_6 generation by the single-seed-descent selfing scheme. Qishu Meshullash was collected from a feral population growing in Israeland was obtained from the Israel Gene Bank. It was identified as subsp. *collosus* based on the recent clarifications of this taxon by (John *et al.*, 2013) and (Endl *et al.*, 2018).

<u>Tam Dew × PI164323 (*HDA192*, F5)</u>: Progenies of the cross between the sweet-fleshed Tam Dew and the non-sweet accession PI164323 (ssp. *melo*, Adzhur Group) were grown in a greenhouse at Newe Ya'ar and ~150 lines were advanced to the F5 generation by single-seed-descent selfing.

<u>Sakata's Sweet × Doya's Faqqous (HDA243, F₃)</u>: The cross between the sweet-fleshed 'Sakata's Sweet' (SAS, ssp. *agrestis*, Makuwa Group) and the locally grown, non-sweet accession which we named 'Doya's Faqqous' (ssp. *melo*, Flexuosus Group) was grown in a greenhouse at Newe Ya'ar and advanced to the F5 generation by single-seed-descent selfing.

<u>PI161375 × Early Silver Line ($HDA232, F_2$)</u>: The cross between the mottled-rind PI161375 (ssp. *agrestis*, Chinensis Group) and the non-mottled-rind <u>Early Silver Line</u> (ESL, ssp. *agrestis*, Makuwa Group) was grown in a greenhouse at Newe Ya'ar and F2 seeds were produced as part of the $HDA25F_2$ seed production.

All the above populations were grown in the open field or in the greenhouse at Newe Ya'ar under common practices, as previously described (Gur *et al.*, 2017; Oren *et al.*, 2020).

Trait evaluation

Accepted Article

<u>Sugars analysis</u>: Five mature fruits per plot were sampled in multiple harvests, with a total of 10-15 fruits per accession being analyzed. Concentrations of total soluble solids (TSS, expressed in $^{\circ}$ Bx) were measured on flesh samples from each of the five fruits separately, using a hand-held refractometer (Atago Co. Ltd., Tokyo, Japan). Approximately one gram of frozen mesocarp tissue, taken from the center-equatorial portion of the fruit, was placed in 80% EtOH and soluble sugars (glucose, fructose and sucrose) were extracted and analyzed by HPLC as previously described (Petreikov *et al.*, 2006).

<u>Characterization of rind color and mottling pattern</u>: Rind color intensity (light vs. dark) and mottling (mottled vs. smooth) were scored visually on young fruits on the vines at ~10-20 days post anthesis, and after harvest of mature fruit.

<u>Response to fusarium wilt</u>: Protocols for pathogen and plant inoculation were described in details by Burger *et al.*, (2003). *Fusarium oxysporum* f.sp. *melonis* used for the inoculation was maintained on yeast extract medium at 27 °C in the dark. Conidial suspensions for seedling inoculation were prepared by macerating 1-week-old cultures with 100 mL water (Cohen *et al.*, 1996). Seeds of the *MelonCore25* accessions were sown in sandy soil. Two days after emergence, the seedlings were removed from the soil and washed under running tap water. The roots were pruned to approximately half their length, and the seedlings were then inoculated by dipping the roots in a conidial suspension (10^6 conidia mL⁻¹) for 2 minutes. The inoculated seedlings were then transplanted into 250 mL pots containing new, disease-free, sandy soil. Each accession was represented by 35 seedlings (five pots × seven plants per pot). The first wilting symptoms were evident 7 days after inoculation. The number of wilted plants was recorded twice a week for two weeks and disease incidence was calculated.

<u>FORC inoculation and symptoms scoring</u>: Pathogen and plant inoculation were described in detail by (Cohen *et al.*, 2015). Briefly, melon seeds were sown in autoclaved sandy soil. Four days post emergence, seedlings were removed from the soil and washed under running tap water. The roots were gently pruned and the seedlings were inoculated by dipping the roots in a suspension of 10⁶ conidia mL⁻¹ for 5 min. The inoculated seedlings were then transplanted into 250 mL pots containing clean sand and maintained in a growth room at a photoperiod regime of 12 h light/darkness. The seedlings were observed for 18 days following inoculation with FORC. The first disease symptom on 'Dulce' plants, leaf chlorosis, appeared six days after inoculation. Subsequently, the susceptible plants exhibited necrosis of the leaf margins, browning of the base of the stem, and wilting of the leaves and apical portion of the stem. The

plant crown then became necrotic and the plant died. Eighteen days after inoculation, each plant was scored as symptomless (resistant) or as dead, wilted, necrotic or chlorotic (susceptible).

<u>Macrophomina phaseolina</u> inoculation and symptom scoring: Pathogen and plant inoculation were described in details by (Cohen *et al.*, 2022). Briefly, *M. phaseolina* (isolate no. NY 198) was isolated from melon plants. This isolate was tested routinely and is highly pathogenic. Melon plants at the age of 5 true leaves were inoculated, approximately 1 cm above ground level, by stabbing them with wooden toothpicks infested with *M. phaseolina* leaving the toothpick in the stab wound (Cohen *et al.*, 2016). Disease severity was scored based on lesion development and size on a 0-to-5 scale, with 0 = symptomless plant, 1 = initial small (< 1 cm) lesion, 2=1–3 cm lesion, 3 = stem rotting of 2–4cm, 4 = stem rotting of >5 cm, 5 = plant dead, completely necrotic. Accessions having average scores of \leq 2.0 were considered to be resistant to *M. phaseolina* whilst those having average scores of \geq 3.0 were considered to be susceptible.

Statistical analyses

Accepted Article

The JMP ver. 14.0.0 statistical package (SAS Institute, Cary, NC, USA) was used for all the general statistical analyses (i.e. Frequency distributions, correlations, analyses of variance and mean comparisons).

Genome-wide association mapping and QTL analysis

Genome-wide association (GWA) analysis of the mottled rind trait on the *GWAS180* panel was performed in TASSEL v.5.2.43 (Bradbury *et al.*, 2007) using the mixed-linear model (MLM) function, as previously described by Gur *et al.*, (2017). Linkage analysis and QTL mapping were performed using R/qtl (Broman *et al.*, 2003) as described in Oren *et al.*(2020).

DNA preparation and long-read Nanopore sequencing

High molecular weight (HMW) DNA extraction and size selection was carried out either using a modified CTAB protocol as previously described in Oren *et al.*, (2022) or using Macherey-Nagel NucleoBond HMW DNA kit (REF 740160, <u>www.mn-net.com</u>). Library preparation was carried out following Oxford Nanopore Technologies (ONT) guidelines as described in Oren *et al.*, (2022) and sequenced on MinION FLO-MIN106D flow cells. Base calling was done using the GPU version of Guppy (V5.0.11, Oxford Nanopore Technologies) using the HAC model.

Genome assembly and scaffolding

ONT reads were assembled using the Flye assembler v2.9 (Kolmogorov *et al.*, 2019), by a previously detailed method (Oren *et al.*, 2022), with the following adjustments: Following initial contig assembly, only one round of ONT-based polishing was carried out using the builtin Flye polishing function, followed by one round of Illumina reads polishing using Pilon v1.24 (Walker *et al.*, 2014). Reference-guided scaffolding of contigs into pseudomolecules was carried out using RagTag v2.1.0 (Alonge *et al.*, 2019) with the 'Harukei-3' melon as a reference (Yano *et al.*, 2020). We used the 'Harukei-3'genome to avoid possible inverted assembly regions on chromosome 6 of DHL92 (Yang *et al.*, 2020).

Characterization of SVs across MelonCore25 and 300 HDA25 crosses

Structural variants (SVs) were characterized by whole-genome alignment of each of the *MelonCore25* genomes to the DHL92 v4.0 genome using nucmer4 (Marçais *et al.*, 2018) with -maxmatch -1 100 - c 500 parameters. The resulting delta file was then analyzed for InDels

with command-line scripts of Assemblytics (v1.2.1) (Nattestad and Schatz, 2016), setting unique sequence length for anchors at 10kb. For short InDels database, InDel size was limited to a range of 15-2000bp—fragment sizes that can be effectively amplified using standard PCR assay. The Assemblytics output tables of all *MelonCore25* lines were concatenated, filtered to InDel-type variants and split by accession to create a matrix of 194K InDels with a separate presence/absence column per accession for each InDel. A simple conditional script was used to list the presence/absence of each InDel across the 300 diallel hybrids obtained from using*MelonCore25*. The final InDel table was composed of 115,802 InDels, their attributes and 325 presence/absence columns (25 accessions and 300 hybrids, **Sup. Table 3**).

Validation of selected InDels and genotyping of MT-2 and Fom-2 by PCR

Twenty-two InDels ranging in size from 16 to 1209 bp that showed polymorphism between DUL and PI414 were selected from our SV database (**Sup. Table 4**), and generated using Assemblytics (v1.2.1) (Nattestad and Schatz, 2016) for validation. Primers were designed using Gene Runner version 6.5. PCR, with an annealing temperature of 56°C, was performed on genomic DNA of Dulce, PI414 and their F1, using 2xPCRBIO HS Taq Mix Red (PCRBIOSYSTEMS, UK). The products were separated on a 2.5% or 1.2% agarose gel for 1-2 hours (**Sup. Figure 2, Sup. Table 4**). Primers for *MT-2* and *FOM-2* markers are listed in **Sup. Table 8**.

Gene models annotation

Article

Accepted A

Draft gene models were annotated using a lift-over approach from two reference sources, the ssp. melo Reticulatus line 'Harukei-3' (Yano et al., 2020) and the ssp. agrestis line HS (Yang et al., 2020) to capture genes from both subspecies. Liftoff v1.6.3 (Shumate and Salzberg, 2021) was run separately for each reference with lenient thresholds on gene coverage and identity (-exclude partial -a 0.5 -s 0.5 -copies), to prevent stringent filtering before determining values that accurately capture the pan and core genes repertoire. Custom filtration is possible on the annotation files based on sequence similarity values in the attributes column of the gff3 file (coverage; extra copy number ; sequence ID). To keep unique gene models from each gff3 file, the gene levels were converted to a bed file using gfftobed (v1.3, https://github.com/jacobbierstedt/gfftobed) and BEDTools intersect was used to mark overlapping gene models lifted over from 'HS' and 'Harukei-3' (Quinlan and Hall, 2010). In the overlapping areas, gene models from the 'HS' were removed as the annotation methods used for the 'Harukei-3' assembly were more rigorous. Prior to merging unique annotations from both sources, an 'HS' prefix was added to all 'HS' gene models, followed by sorting and tidying. Filtering, naming and sorting of the gene-models was done with AGAT toolkit (Version v0.8.0, https://www.doi.org/10.5281/zenodo.3552717) and tidying was done with genometools gff3 tool (v.1.6.2, https://github.com/genometools/genometools).

DATA AVAILABILITY STATEMENT

All the data supporting the findings of this study are available within the paper and within its supplementary materials published online. Raw sequences and FASTA files of genome assemblies can be found at NCBI BioProject PRJNA726743. Gene annotation files for each of the genomes (gff format) are available at Zenodo repository https://doi.org/10.5281/zenodo.7212598.

ACKNOWLEDGMENTS

We acknowledge Dr. Harry Paris for fruitful discussions and critical editing of the manuscript. We thank Fabian Baumkoler and the Newe-Yaar farm team for technical assistance in setting the field trials and for plant maintenance. We thank Yelena Yeselson for technical assistance in sugar analyses. Funding for this research was provided by the United States-Israel Binational Agricultural Research and Development Fund (BARD) grants no. IS-4911-16 and IS-5284-20.

AUTHOR CONTRIBUTIONS

AG conceived the research plan; EO, AD and AG designed the experiments; US, SO, AD, JB and AG were responsible for development of plant genetic materials; AD, EO, GT, TI, AM, IH and AG performed the field experiments and collected the data; AD was responsible for digital data collection; ME, SO and RC performed phytopathology screens; YT and AAS contributed to sugars analyses; GT performed DNA markers analyses. EO, CR, TL and ESB planned and performed the genomic sequencing and *de novo* assembly pipeline. EO analyzed the genomic data and assembled the genomes. EO and AG analyzed the results. AG and EO wrote the manuscript. All authors discussed the results, reviewed and approved the final version of the manuscript.

CONFLICT OF INTEREST

Accepted Article

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Supplementary Figure 1: Workflow for using the pan-genome and multi-parental mapping framework.

Supplementary Figure 2: PCR validation of 22 InDels. InDel numbers correspond to those in Sup. Table 4.

Supplementary Figure 3: Mode of inheritance of TSS across the *HDA20* set: 190 hybrids and their parents.

Supplementary Figure 4: Variation in disease severity index (DSI) to *Macrophomina phaseolina* across *MelonCore25.* **a**) Frequency distribution of DSI across *MelonCore25.* **b**) Projection of DSI on the genetic PCA. Diverse crosses are indicated with dashed lines and hybrid code (HDA number).

Supplementary Figure 5: Comparisons between ONT and Ilumina reads alignments across 8 accessions carrying the FOM 2 insertion. Purple boxes within ONT reads represent the \sim 1,100 bp insertions.

Supplementary Table 1: Nanopore long-read sequencing statistics and genetic PCA coordinates of MelonCore25.

Supplementary Table 2: Large SVs across *MelonCore25*.

365313x, ja, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pj.16021 by Cornell University, Wiley Online Library on [2//11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Supplementary Table 3: Short InDels across MelonCore25 and 300 HDA25 crosses.

Supplementary Table 4: List of 22 InDels validated by PCR.

Supple mentary Table 5: SNPs and short exonic polymorphisms within genes in the mottledrind interval (*MT-2*) on chromosome 2, across *MelonCore25*.

Supplementary Table 6: SVs within the mottled-rind interval (*MT-2*) on chromosome 2, across *MelonCore25*.

Supplementary Table 7: Exonic polymorphisms in the MELO3C021831 *FOM-2* gene, across *MelonCore25*.

Supplementary Table 8: Primers for the *MT-2* and *FOM-2* markers.

REFERENCES

Accepted Article

- Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F., Lippman, Z. and Schatz, M. (2019) RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol.*, 20, 519637.
- Argyris, J.M., Díaz, A., Ruggieri, V., et al. (2017) QTL analyses in multiple populations employed for the fine mapping and identification of candidate genes at a locus affecting sugar accumulation in melon (*Cucumis melo L.*). Front. Plant Sci., 8, 1–20.
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, 3, 1–7.
- Bayer, P.E., Golicz, A.A., Scheben, A., Batley, J. and Edwards, D. (2020) Plant pangenomes are the new reference. *Nat. Plants*, 6.
- Biémont, C. (2008) Genome size evolution: Within-species variation in genome size. *Heredity (Edinb).*, 101, 297–298.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23, 2633–2635.
- Broman, K.W., Wu, H., Sen, S. and Churchill, G.A. (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19**, 889–890.
- Burger, Y., Katzir, N., Tzuri, G., Portnoy, V., Saar, U., Shriber, S., Perl-Treves, R. and Cohen, R. (2003) Variation in the response of melon genotypes to *Fusarium oxysporum* f. sp. melonis race 1 determined by inoculation. *Plant Pathol.*, **52**, 204–211.
- Burger, Y., Paris, H.S., Cohen, R., Katzir, N., Tadmor, Y., Lewinsohn, E. and Schaffer, A.A. (2009) Genetic diversity of *Cucumis melo. Hortic. Rev. (Am. Soc. Hortic. Sci).*, 36, 165–198.
- Burger, Y., Sa'ar, U., Paris, H.S., Lewinsohn, E., Katzir, N., Tadmor, Y. and Schaffer, A.A. (2006) Genetic variability for valuable fruit quality traits in *Cucumis melo*. Isr. J. Plant Sci., 54, 233–242.
- Burger, Y. and Schaffer, A.A. (2007) The contribution of sucrose metabolism enzymes to sucrose accumulation in *Cucumis melo. J. Am. Soc. Hortic. Sci.*, **132**, 704–712.

Castanera, R., Ruggieri, V., Pujol, M. and Garcia-mas, J. (2020) An improved melon

- Cohen, R., Blaier, B., Schaffer, A.A. and Katan, J. (1996) Effect of acetochlor treatment on Fusarium wilt and sugar content in melon seedlings. *Eur. J. Plant Pathol.*, **102**, 45–50.
- Cohen, R., Elkabetz, M. and Edelstein, M. (2016) Variation in the responses of melon and watermelon to *Macrophomina phaseolina*. *Crop Prot.*, **85**, 46–51.
- **Cohen, R., Elkabetz, M., Paris, H.S., Freeman, S. and Gur, A.** (2022) Charcoal rot (*Macrophomina phaseolina*) across melon diversity : evaluating the interaction between the pathogen , plant age and environmental conditions as a step towards breeding for resistance. *Eur. J. Plant Pathol.*, **163**, 601–613.
- **Cohen, R., Orgil, G., Burger, Y., et al.** (2015) Differences in the responses of melon accessions to fusarium root and stem rot and their colonization by *Fusarium oxysporum* f. sp. radicis-cucumerinum. *Plant Pathol.*, **64**, 655–663.
- Cohen, S., Itkin, M., Yeselson, Y., et al. (2014) The PH gene determines fruit acidity and contributes to the evolution of sweet melons. *Nat. Commun.*, **5**, 4026.
- Coletta, R. Della, Qiu, Y., Ou, S., Hufford, M.B. and Hirsch, C.N. (2021) How the pangenome is changing crop genomics and improvement. *Genome Biol.*, **22**, 1–19.
- Cuevas, H.E., Staub, J.E., Simon, P.W. and Zalapa, J.E. (2009) A consensus linkage map identifies genomic regions controlling fruit maturity and beta-carotene-associated flesh color in melon (*Cucumis melo* L.). *Theor. Appl. Genet.*, **119**, 741–756.
- Dafna, A., Halperin, I., Oren, E., et al. (2021) Underground heterosis for yield improvement in melon. J. Exp. Bot., 72, 6205–6218.

Accepted Articl

- Dai, N., Cohen, S., Portnoy, V., et al. (2011) Metabolism of soluble sugars in developing melon fruit: a global transcriptional view of the metabolic transition to sucrose accumulation. *Plant Mol. Biol.*, 76, 1–18.
- Diaz, A., Fergany, M., Formisano, G., et al. (2011) A consensus linkage map for molecular markers and quantitative trait loci associated with economically important traits in melon (*Cucumis melo* L.). *BMC Plant Biol.*, **11**, 111.
- **Doudna, J.A. and Charpentier, E.** (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science.*, **346**.
- Elkabetz, M., Paris, H.S., Burger, Y., Hanan, A. and Cohen, R. (2016) Two genes for resistance to *Fusarium oxysporum* f. sp. radicis-cucumerinum in melon (Cucumis melo, Cucurbitaceae). *Sci. Hortic. (Amsterdam).*, 201, 57–60.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J. a., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6, 1–10.
- Endl, J., Achigan-Dako, E.G., Pandey, A.K., Monforte, A.J., Pico, B. and Schaefer, H. (2018) Repeated domestication of melon (*Cucumis melo*) in Africa and Asia and a new close relative from India. *Am. J. Bot.*, 105, 1662–1671.
- Feder, A., Burger, J., Gao, S., et al. (2015) A Kelch domain-containing F-box coding gene negatively regulates flavonoid accumulation in *Cucumis melo* L. *Plant Physiol.*, 169, 1714–1726.
- Galpaz, N., Gonda, I., Shem-Tov, D., et al. (2018) Deciphering genetic factors that determine melon fruit-quality traits using RNA-Seq-based high-resolution QTL and eQTL mapping. *Plant J.*, 94, 169–191.

- Gao, L., Gonda, I., Sun, H., et al. (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.*, **51**, 1044–1051.
- Garcia-Mas, J., Benjak, a., Sanseverino, W., et al. (2012) The genome of melon (*Cucumis melo* L.). Proc. Natl. Acad. Sci., 109, 11872–11877.
- Gur, A., Tzuri, G., Meir, A., et al. (2017) Genome-wide linkage-disequilibrium mapping to the candidate gene level in melon (*Cucumis melo*). Sci. Rep., 7, 9770.
- Harel-Beja, R., Tzuri, G., Portnoy, V., et al. (2010) A genetic map of melon highly enriched with fruit quality QTLs and EST markers, including sugar and carotenoid metabolism genes. *Theor. Appl. Genet.*, **121**, 511–533.
- Ho, S.S., Urban, A.E. and Mills, R.E. (2020) Structural variation in the sequencing era. *Nat. Rev. Genet.*, **21**, 171–189.
- Jayakodi, M., Padmarasu, S., Haberer, G., et al. (2020) The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature*, **588**, 284–289.
- John, K.J., Scariah, S., Nissar, V.A.M., Latha, M., Gopalakrishnan, S., Yadav, S.R. and Bhat, K. V (2013) On the occurrence, distribution, taxonomy and genepool relationship of Cucumis callosus (Rottler) Cogn., the wild progenitor of *Cucumis melo* L. from India. *Genet. Resour. Crop Evol.*, **60**, 1037–1046.
- Joobeur, T., King, J.J., Nolin, S.J., Thomas, C.E. and Dean, R.A. (2004) The fusarium wilt resistance locus Fom-2 of melon contains a single resistance gene with complex features. *Plant J.*, **39**, 283–297.
- Kolmogorov, M., Yuan, J., Lin, Y. and Pevzner, P.A. (2019) Assembly of long, errorprone reads using repeat graphs. *Nat. Biotechnol.*, 37, 540–546.

Accepted Article

- Leida, C., Moser, C., Esteras, C., Sulpice, R., Lunn, J.E., Langen, F. de, Monforte, A.J. and Picó, B. (2015) Variability of candidate genes, genetic structure and association with sugar accumulation and climacteric behavior in a broad germplasm collection of melon (*Cucumis melo* L.). *BMC Genet.*, 16, 1–17.
- Li, H., Wang, S., Chai, S., et al. (2022) Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nat. Commun.*, **13**, 1–14.
- Ling, J., Xie, X., Gu, X., Zhao, J., Ping, X., Li, Y., Yang, Y., Mao, Z. and Xie, B. (2021) High-quality chromosome-level genomes of *Cucumis metuliferus* and *Cucumis melo* provide insight into Cucumis genome evolution. *Plant J.*, 107, 136–148.
- Liu, S., Gao, P., Zhu, Q., Zhu, Z., Liu, H., Wang, X., Weng, Y., Gao, M. and Luan, F. (2020) Resequencing of 297 melon accessions reveals the genomic history of improvement and loci related to fruit traits in melon. *Plant Biotechnol. J.*, 18, 2545– 2558.
- Liu, Y., Du, H., Li, P., et al. (2020) Pan-Genome of Wild and Cultivated Soybeans. *Cell*, 182, 162-176.e13.
- Lv, J., Fu, Q., Lai, Y., Zhou, M. and Wang, H. (2018) Inheritance and gene mapping of spotted to non-spotted trait gene CmSp-1 in melon (*Cucumis melo* L. var. chinensis Pangalo). *Mol. Breed.*, 38, 1–9.
- Ma, L., Wang, Q., Zheng, Y., et al. (2022) Cucurbitaceae genome evolution, gene function, and molecular breeding. *Hortic. Res.*, 9.
- Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L. and Zimin, A. (2018) MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.*,

14, 1–14.

Accepted Article

- Martínez-Martínez, C., Gonzalo, M.J., Sipowicz, P., et al. (2022) A cryptic variation in a member of the Ovate Family Proteins is underlying the melon fruit shape QTL fsqs8.1. *Theor. Appl. Genet.*, **135**, 785–801.
- Moing, A., William Allwood, J., Aharoni, A., et al. (2020) Comparative metabolomics and molecular phylogenetics of melon (*Cucumis melo*, cucurbitaceae) biodiversity. *Metabolites*, 10.
- Monforte, A.J., Diaz, A., Caño-Delgado, A. and Knaap, E. Van Der (2014) The genetic basis of fruit morphology in horticultural crops: Lessons from tomato and melon. *J. Exp. Bot.*, **65**, 4625–4637.
- Morrell, P.L., Buckler, E.S. and Ross-Ibarra, J. (2011) Crop genomics: advances and applications. *Nat. Rev. Genet.*, **13**, 85–96.
- Nattestad, M. and Schatz, M.C. (2016) Assemblytics: A web analytics tool for the detection of variants from an assembly. *Bioinformatics*, **32**, 3021–3023.
- **Oren, E., Tzuri, G., Dafna, A., et al.** (2022) QTL mapping and genomic analyses of earliness and fruit ripening traits in a melon recombinant inbred lines population supported by de novo assembly of their parental genomes . *Hortic. Res.*, **9**.
- Oren, E., Tzuri, G., Dafna, A., et al. (2020) High-density NGS-based map construction and genetic dissection of fruit shape and rind netting in Cucumis melo. *Theor. Appl. Genet.*, 133, 1927–1945.
- Oren, E., Tzuri, G., Vexler, L., et al. (2019) Multi-allelic APRR2 gene is associated with fruit pigment accumulation in melon and watermelon. J. Exp. Bot., 70, 3781–3794.
- **Oumouloud**, A. and Otmani, M. El (2015) Molecular characterization of Fom 1 gene and development of functional markers for molecular breeding of resistance to Fusarium race 2 in melon. , 491–501.
- Paris, H.S., Lebeda, A., Křistkova, E., Andres, T.C. and Nee, M.H. (2012) Parallel evolution under domestication and phenotypic differentiation of the cultivated subspecies of *cucurbita pepo* (Cucurbitaceae). *Econ. Bot.*, 66, 71–90.
- Pereira, L., Ruggieri, V., Pérez, S., Alexiou, K.G., Fernández, M., Jahrmann, T. and Pujol, M. (2018) QTL mapping of melon fruit quality traits using a high-density GBSbased genetic map. *BMC Plant Biol.*, 18, 1–17.
- Pereira, L., Santo Domingo, M., Ruggieri, V., et al. (2020) Genetic dissection of climacteric fruit ripening in a melon population segregating for ripening behavior. *Hortic. Res.*, 7, 187.
- Petreikov, M., Shen, S., Yeselson, Y., Levin, I., Bar, M. and Schaffer, A.A. (2006) Temporally extended gene expression of the ADP-Glc pyrophosphorylase large subunit (AgpL1) leads to increased enzyme activity in developing tomato fruit. *Planta*, **224**, 1465–1479.
- Pichot, C., Djari, A., Tran, J., et al. (2022) Cantaloupe melon genome reveals 3D chromatin features and structural relationship with the ancestral cucurbitaceae karyotype. *iScience*, 25, 103696.
- Pitrat, M., Hanelt, P. and Hammer, K. (2000) Some comments on infraspecific classification on cultivars of melon. In N. Katzir and H. S. Paris, eds. *Proceedings of Cucurbitaceae 2000, 7th Eucarpia Meeting on Cucurbit Genetics and Breeding. Acta Horticulturae.* pp. 29–36.

- Punja, Z.K. and Parker, M. (2000) Development of fusarium root and stem rot, a new disease on greenhouse cucumber in British Columbia, caused by *Fusarium oxysporum* f. sp. radicis-cucumerinum. *Can. J. Plant Pathol.*, 22, 349–363.
- Purugganan, M.D. and Jackson, S.A. (2021) Advancing crop genomics from lab to field. Nat. Genet., 53, 595–601.
- Qin, P., Lu, H., Du, H., et al. (2021) Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell*, **184**, 3542-3558.e16.
- Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842.
- Ríos, P., Argyris, J., Vegas, J., et al. (2017) ETHQV6.3 is involved in melon climacteric fruit ripening and is encoded by a NAC domain transcription factor. *Plant J.*, 91, 671– 683.
- Schaffer, A.A., Aloni, B. and Fogelman, E. (1987) Sucrose metabolism and accumulation in developing fruit of *Cucumis*. *Phytochemistry*, 26, 1883–1887.
- Shen, J., Xu, X., Zhang, Y., Niu, X. and Shou, W. (2021) Genetic mapping and identification of the candidate genes for mottled rind in *Cucumis melo* L. *Front. Plant Sci.*, 12, 1–9.
- Shumate, A. and Salzberg, S.L. (2021) Liftoff: accurate mapping of gene annotations. *Bioinformatics*, **37**, 1639–1643.
- Song, J.M., Guan, Z., Hu, J., et al. (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants*, 6.
- Tzuri, G., Zhou, X., Chayut, N., et al. (2015) A "golden" SNP in CmOr governs fruit flesh color of melon (*Cucumis melo*). *Plant J.*, 82, 267–279.

Accepted Article

- Vakalounakis, D.J., Doulis, A.G. and Klironomou, E. (2005) Characterization of *Fusarium* oxysporum f. sp. radicis-cucumerinum attacking melon under natural conditions in Greece. *Plant Pathol.*, 54, 339–346.
- Walker, B.J., Abeel, T., Shea, T., et al. (2014) Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement J. Wang, ed. *PLoS One*, 9, e112963.
- Wang, S., Yang, J. and Zhang, M. (2011) Developments of functional markers for Fom 2 mediated fusarium wilt resistance based on single nucleotide polymorphism in melon (*Cucumis melo* L.). *Mol. Breed.*, 27, 385–393.
- Yang, J., Deng, G., Lian, J., Garraway, J., Niu, Y., Hu, Z., Yu, J. and Zhang, M. (2020) The chromosome-scale genome of melon dissects genetic architecture of important agronomic traits. *iScience*, 23, 101422.
- Yano, R., Ariizumi, T., Nonaka, S., Kawazu, Y., Zhong, S., Mueller, L., Giovannoni, J.J., Rose, J.K.C. and Ezura, H. (2020) Comparative genomics of muskmelon reveals a potential role for retrotransposons in the modification of gene expression. *Commun. Biol.*, 3, 1–13.
- Zhang, H., Li, X., Yu, H., et al. (2019) A high-quality melon genome assembly provides insights into genetic basis of fruit trait improvement. *iScience*, 22, 16–27.
- Zhao, G., Lian, Q., Zhang, Z., et al. (2019) A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nat. Genet.*, 51, 1607–1615.

Figure Legends

Figure 1: The *MelonCore25* set and mapping platform development scheme. a) Mature fruits of the 25 founders, ordered by subspecies and horticultural (cultivar-) groups. b) Schematic diagram for the components and development process of the multi-parental framework.

Figure 2: Melon pan-genome and distribution of structural variations (SVs) across *HDA25.* **a)** Phylogenetic tree of the 25 founder lines presented with variation in genome size and examples of large inter-chromosomal inversions on chromosomes 1 and 11 in ssp. *agrestis.* **b)** The genomes of the ssp. *agrestis* accessions are significantly smaller than those of the ssp. *melo* accessions. **c)** Size distribution of 108,011 large SVs (>2 Kb). The X axis is presented in log scale. **d)** Distribution of allele frequency of 108,011 large SVs (>2 Kb). **e)** Size distribution of 86,669 short (<2 Kb) bi-allelic InDels across MelonCore25. X axis is presented in log scale. **f)** Distribution of allele frequency across 86,669 bi-allelic sites (<2 Kb). **g)** Pan InDels analysis. Horizontal black lines are the mean number of pan InDels under corresponding number of genomes. Dark boxes are the 50th percentile and light boxes are 99th percentile **h)** Correlation between parental distance (calculated using 24,000 genome-wide GBS SNPs) and number of polymorphic short InDels across the 300 *HDA25* crosses. **i)** Number of polymorphic short InDels by chromosomes and bins, in the pan-genome and two crosses. In parentheses is the total number of short Indels for each group.

Figure 3: Genetic analysis and mapping of mottled rind. a) Representative mottled rind fruits from the *GWAS180* collection. b) Parallel frequencies of mottled- and non-mottled-rind accessions in the *GWAS180* collection and the *MelonCore25* set. c) Projection of rind pattern (mottled or non-mottled) on the genetic PCA plot of the *MelonCore25* set. Dashed lines indicate crosses (and derived segregating populations) used for genetic analysis of the mottled-rind trait. d) PI414723 and DUL, the parents of the RILs population used for mapping the mottled-rind trait. e) Linkage mapping results of the *MT-2* locus to chromosome 2. f) Zoom in on chromosome 2 mapping in the RILs of the cross PI414723 × DUL and comparison with a Manhattan plot derived from GWA analysis of the *GWAS180* collection. g) Substitution-mapping validation for the *MT-2* locus using recombinants from the RILs of the cross PI414723×DUL. Green cells represent flanking markers for the interval. The yellow cell is the rind phenotype column. A = DUL allele, B = PI414723 allele. The GWAS peak interval is represented by the red horizontal line.

Figure 4: Multi-parental characterization of the mottled rind trait. a) Inheritance of rind pattern across the HDA20 set (190 half-diallel F1s and their 20 parents). Each cell represents the combination of parents from the horizontal and vertical axes. Squares arranged in the diagonal represent the phenotypes of the parents. **b)** Contingency analysis for the segregation of rind pattern against the MT-2 marker in the F₂ of the cross QME×TAD. Het=Heterozygote. **c)** Contingency analysis for the segregation of rind pattern against the MT-2 marker in the F₂ of the cross PI161375×ESL. Het=Heterozygote. Examples for young fruit segregants (mottled and non-mottled) are displayed. **d)** Contingency analysis for the segregation of rind color against the APRR2 marker in the F₂ of the cross PI161375×ESL. Het=Heterozygote. Examples for the segregation of rind color against the APRR2 marker in the F₂ of the cross PI161375×ESL. Het=Heterozygote. Examples for the segregation of rind color against the APRR2 marker in the F₂ of the cross PI161375×ESL. Het=Heterozygote. Examples for the segregation of rind color against the APRR2 marker in the F₂ of the cross PI161375×ESL. Het=Heterozygote. Examples for the segregation of rind color against the APRR2 marker in the F₂ of the cross PI161375×ESL. Het=Heterozygote. Examples for mature fruit segregants (light and dark rind) are displayed. **e)** Contingency analysis for the segregation of rind pattern against the MT-2 marker in the RILs from the cross PI414723×DUL. **f)** Schematic pivot table that presents the co-segregation of the APRR2 and MT-2 markers

against rind color and pattern in the F_2 of the cross PI161375×ESL. **g**) Allelism test for the mottled-rind trait in the F_2 of the cross DUD×PSR. **h**) Allelism test for the mottled rind trait in the F_2 of the cross QME×PSR. **i**) Allelism test for the mottled rind trait in the F_2 of the cross QME×PSR. **i**) Allelism test for the mottled rind trait in the F_2 of the cross QME×PI414723.

Figure 5: Variation in fruit sugars across *MelonCore25* and selected segregating populations for genetic mapping. a) Distribution of TSS of mature fruits across *MelonCore25*. b) Projection of accession TSS on the genetic PCA. Dashed lines represent crosses (and derived segregating populations) used for genetic dissection of fruit sugar accumulation. c) Correlation between content of the monosaccharides, glucose and fructose, and disaccharide, Sucrose, in mature fruits of *MelonCore25*. d) Selection of tails segregants through the analysis of correlation between F_3 and F_4 lines of the cross TAD×QME (*HDA008*). High and low TSS tails are represented in red and blue, respectively. e) Selection of tail segregants through the analysis of correlation between F_4 and F_5 lines of the cross TAD×PI164323 (*HDA192*). High and low TSS tails are represented in red and blue, respectively. f) Selection of tail segregants through the F_4 population from the cross SAS×DOYA (*HDA243*). High and low TSS tails are represented in red and blue, respectively. f) Selection of tail segregants through the regulation from the cross SAS×DOYA (*HDA243*). High and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red and blue, respectively. high and low TSS tails are represented in red

Figure 6: Variation in response to artificial inoculation with *Fusarium oxysporum* f.sp. *melonis* races 1 and 2 and haplotypic variation in the *FOM-2* gene across *MelonCore25*. a) Projection of accession response to inoculation with Fusarium race 1 on the genetic PCA. b) Projection of accession response to inoculation with Fusarium race 2 on the genetic PCA. c) Haplotypic variation in the *FOM-2* gene (MELO3C021831) across *MelonCore25*. Vertical colored lines are SNPs. Gray histograms reflect short read depth. The dashed rectangle in haplotype 3 is the1,100 bp insertion discovered through *Nanopore* sequencing and *de novo* assemblies. Remarkably, a ~500 bp deletion was initially indicated in this region based on absence of Illumina short read alignments. The actual insertion was discovered only based on the long-read assemblies (**Sup. Figure 5**). Colored dots correspond to the Fusarium race 1 response as presented in **Figure 6a**. Part of the gene model is presented below the haplotypic view. Triangles bellow the gene model are PCR primers for FOM-2 CAPs markers. Blue: marker from (Oumouloud and Otmani, 2015). Red: CAPs marker developed for the Newe Ya'ar breeding program (**Sup. Table 8**). d) Projection of accession *FOM-2* haplotype on the genetic PCA. e) Gel image of PCR validation of the 1,100 bp InDel.

Accepted Article

Figure 7: Variation in response to *Fusarium oxysporum* f. sp. *radicis-cucumerinum* (FORC) across *MelonCore25* and QTL mapping of TAD×DUL RILs. a) Experimental design and examples of disease response of susceptible (S) and resistant (R) accessions, 7–14 days post inoculation. b) Projection of the responses of the 25 accessions to FORC on the genetic PCA. The dashed line marks a cross, HDA034, used for the QTL mapping. c) Frequency distribution of FORC Disease Symptoms Index (DSI) across 153 TAD × DUL RILs. Responses of parents are marked. d) Manhattan plot for linkage mapping of response to FORC in the TAD × DUL RILs. e) Allelic effect of the chromosome 7 QTL. For both respective genotypic groups, DUL and TAD, the two gray boxes represent the standard deviation and the blank spaces in-between the allelic mean.

Table 1: List and taxonomic affiliation of the 25 founders of *MelonCore25* and their genome assembly statistics.

Table 1: List of 25 founders (MelonCore25) and genome assembly statistics

Line Short Name	Variety Name	Sub-Species	varietyType	Assembly size (Mb)	Total Number of Contigs	Total Number of Scaffolds	Largest Contig (Mb)	Contigs size N50 (Mb)	Contigs size N90 (Mb)	Assembly Complete BUSCOs (% of 2325 genes)	LAI	Number of Annotated genes
ARJ	Arka Jeet	melo	Khandalak	364.95	525	327	14.29	4.90	1.48	97.94%	13.22	36773
AY	Ananas Yoqne'am	melo	Ameri	367.46	280	176	22.86	8.18	2.25	97.85%	13.28	37183
BAHC	Bately's Hermaphrodite	agrestis	Chinensis	361.94	247	170	22.73	11.24	3.19	97.94%	12.19	36981
BDR	Branco De Ribatejo	melo	Inodorus	365.98	462	346	24.52	7.92	2.85	97.85%	13.20	37136
BEL	Bellegarde	melo	Cantalupensis	373.83	226	142	23.52	8.03	3.09	97.76%	14.15	37193
DOYA	Doya's Faqqous	melo	Flexuosus	366.72	473	277	21.03	5.45	1.72	97.72%	13.05	36513
DUD	Duda'im 3	melo	Duda'im	362.85	357	214	16.18	5.78	1.78	97.89%	11.90	36602
DUL	Dulce	melo	Reticulatus	365.53	124	63	22.65	10.68	3.17	97.63%	11.87	36175
ESL	Early Silver Line	agrestis	Makuwa	358.58	560	163	8.45	1.94	0.48	97.94%	11.11	36345
INB	Indian Best	melo	Khandalak	363.58	231	117	24.99	6.04	2.16	97.89%	12.88	36626
KRY	Krimka	melo	Reticulatus	368.44	441	295	34.51	9.72	1.58	97.76%	13.23	37158
NA	Noy 'Amid	melo	Inodorus	367.29	239	155	27.05	12.63	3.11	97.85%	13.98	37259
NDD1	NDD1	melo	Cantalupensis	365.12	282	208	40.92	10.69	2.99	97.85%	14.30	37122
NY	Noy Yizre'el	melo	Cantalupensis	365.69	221	148	23.95	12.55	2.96	97.85%	13.48	36919
OHG	Ohgon	agrestis	Makuwa	360.67	174	113	23.11	13.81	3.53	97.94%	12.36	36883
PI161375	PI161375	agrestis	Chinensis	360.06	458	242	15.99	4.29	0.99	97.85%	11.65	36593
PI164323	PI164323	melo	Adzhur	367.92	1029	750	22.87	5.69	0.81	98.02%	12.58	36394
PI414723	PI414723	agrestis	Momordica	363.40	230	157	22.82	11.89	3.30	98.06%	12.80	36458
PSR	Piel De Sapo Redon	melo	Inodorus	368.71	258	165	26.60	9.26	2.91	97.89%	13.40	37232
QME	Qishu Meshullash	C. collosus	feral	363.55	269	184	32.87	15.49	2.90	98.02%	13.10	36578
SAS	Sakata's Sweet	agrestis	Makuwa	361.03	432	309	20.76	6.75	2.27	97.94%	11.74	36725
TAD	Tam Dew	melo	Inodorus	364.88	173	98	23.60	12.13	3.14	97.94%	13.33	37120
TOG	Tokyo Giant	agrestis	Conomon	361.17	273	161	21.63	9.11	1.73	97.94%	12.26	36802
тит	Tendral Verde Tardio	melo	Inodorus	364.82	245	142	42.25	7.38	2.44	97.89%	13.45	36970
VEP	Védrantais	melo	Cantalupensis	365.38	528	392	22.55	6.88	1.60	97.81%	13.33	36984
DHL92v4.0 Harukei-3		melo	Reticulatus	357.74 372.24	1192 138	1192 138	5.66 23.54	0.69 8.25	0.19 1.95	97.55% 97.76%	12.99 13.51	28,299 37,254

Figure 1: MelonCore25 set and mapping platform development scheme

а

Cucumis melo ssp. melo



Ribatejo (BDR)



Védrantais

(VEP)

Cantalupensis

Khandalak



(TAD)



Noy Yizre'el (NY)

Bellegarde (BEL)

Duda'im



Duda'im (DUD)



(DUD)



Indian Best (INB)



Ohgon

(OHG)





Sakata's Sweet (SAS)

Conomon





Tokyo Giant (TOG)



Early Silver

Line (ESL)



NDD1

Redon (PSR)

Adzhur

PI 164323

PI 161375

Chinensis

Batley's

Hermaphrodite

(BAHC)

Ananas Yoqne'am (AY)

Ameri

Reticulatus



Flexuosus

Doya's Faqqous

(DOYA)

Krymka (KRY) Dulce (DUL)

C. melo ssp. Collosus



Qishu Meshullash (QME)



Figure 2: Melon pan-genome and distribution of SVs across HDA25



Pan-genome (115,802) DOYAxQME (45,537) OHGxSAS (5,082)

Figure 3: Genetic analysis and mapping of mottled rind



Figure 4: Multi-parental characterization of mottled rind trait





Figure 6: Variation in response to *Fusarium oxysporum races 1* and 2 and haplotypic variation in FOM-2 gene across *MelonCore25*



Figure 7: variation in response to Fusarium oxysporum f. sp. radicis-cucumerinum (FORC) across MelonCore25 and QTL mapping

