Genomics-aided Development and Characterization of *Cucumis hystrix* Introgression Lines in Cucumber

by

Paradee Thammapichai

A dissertation submitted in partial fulfillment of

the requirements for degree of

Doctor of Philosophy

(Plant Breeding and Plant Genetics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2019

Date of final oral examination: 04/25/2019

The dissertation is approved by the following members of the Final Oral Committee:
Yiqun Weng, Associate Professor, Horticulture
Michael Havey, Professor, Horticulture
Philipp Simon, Professor, Horticulture
Juan Zalapa, Associate Professor, Horticulture
Shawn Kaeppler, Professor, Agronomy

Genomics-aided Development and Characterization of *Cucumis hystrix* Introgression Lines in Cucumber

Paradee Thammapichai

Under the supervision of Professor Yiqun Weng

at the University of Wisconsin-Madison

Abstract

Cucumis hystrix ($2n = 2x = 24$) is the only known species in the genus *Cucumis* that is cross-compatible with cucumber (*C. sativus*, $2n = 2x = 14$) and has potentials for cucumber improvement. To facilitate introgression of *C. hystrix* chromatins into cucumber genetic background, we developed a draft genome assembly for *C. hystrix* (TH1 accession) which contained 16,865 scaffolds (~78× coverage) for a total of 226.0 Mb, representing ~51% of the estimated 447 Mb *C. hystrix* genome. Through genotyping-by-sequencing (GBS), a linkage map for *C. hystrix* was constructed with 1,692 SNP loci, which was then integrated with a previously developed genetic map with 410 SSR markers. The resulting consensus map consisted of 12 linkage groups spanning 1119 cM and anchor 1,321 scaffolds, accounting for 60.2 Mb or ~26.62% of *C. hystrix* draft genome assembly. A karyotype for the *C. hystrix* genome was developed with molecular cytogenetic landmarks which allowed for a better understanding of the syntenic relationships among *C. hystrix*, cucumber and melon (*C. melo*, $2n = 2x = 24$) chromosomes as well as the mechanisms in which the seven cucumber chromosomes were evolved from an x=12 ancestor through dysploid chromosome reduction. *C. hystrix* introgression lines (IL) in elite cucumber backgrounds were developed through marker-assisted backcrossing.

Overall, 334 ILs were identified by SSR markers to carry *C. hystrix* alleles. These ILs included different introgression segments that covered about 57.29% of *C. hystrix* map. With genotyping-by-sequencing (GBS), the introgressions in the 55 ILs included 44 syntenic blocks or about 74.57% coverage of *C. hystrix* genetic map. Each IL has an average 5.4 introgression segments, averaging 1.7 Mbp in size. Homoeologous recombination reduction was observed genome-wide but recombination hotspots in *C. hystrix* were also identified. Phenotypic characterization for resistance to angular leaf spot (ALS) and downy mildew (DM) pathogens was conducted in selected ILs. Both genotypic data and phenotypic data indicated that the intermediate resistance to ALS in nine ILs and the high resistance to DM in four ILs was of *C. hystrix* origin. These *C. hystrix* introgression lines, although time consuming in development, should be valuable resources for cucumber improvement.

Acknowledgements

First and foremost, I would like to thank Dr. Yiqun Weng for the opportunity to be working in this lab. I truly appreciate all the advices and patience he has given me even when I struggled during the past five years. Thanks for being so understanding. The project would not be accomplished without his unlimited support.

I would like to thank Dr. Michael Havey, Dr. Philipp Simon, Dr. Jiming Jiang, Dr. Juan Zalapa, and Dr. Shawn Kaeppler for their willingness to serve on my committee and all the recommendations for my research project.

I would like to give a special thanks to Dr. Junsong Pan whose preliminary work laid a solid foundation for this project and to Dr. Douglas Senalik who gave me a ton of bioinformatic support, especially for development of the *C. hystrix* genome assembly and analysis of GBS data. Along with this, thanks to my IT friend, Dr. Chinawat Isradisaikul, who answered all my questions about writing scripts. I would like to thank Dr. Anchittha Satjarak for guidance in genome annotation. I would like to thank Kristin Haider, Dr. Luming Yang, all members of Weng lab, and Holly Ruess for their help in the lab and in the field, and for their help with many questions or tasks I bombarded them with.

I would like to thank Ms. Kathleen Reitsma at the North Central Plant Introduction Station in Ames, IA for allowing me to use the facility to help with seed increase for my project. I would like to thank Dr. Yonghua Han and Dr. Li He for their help on cytogenetic work, and Dr. Sarah

Patterson to the use of microscope to take pictures of pollen staining. I thank Dr. Todd C. Wehner for DM field data collection in Raleigh, NC during summer 2018. I thank Dr. Yuhui Wang for providing *P. cubensis* isolates and the DM-resistant near-isogenic lines and for help with DM screening and Dr. Alyson Thornton, Harris Moran Seed Company (Sun Prairie, WI) for *P.* syringae isolates and for help with ALS screening process.

I would like to thank my fellow and former graduate Thai students, especially Dr. Anchittha Satjarak. We supported each other during this journey and make Madison feel more like home to me. Most important of all, I would like to thank my Mom and Dad for all the laughs and support they have given me all these years.

I gratefully acknowledge the financial support for my research project by Agriculture and Food Research Initiative Competitive Grant 2013-67013-21105 from the U.S. Department of Agriculture National Institute of Food to Dr. Yiqun Weng and the Royal Thai Government for my graduate study.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1: General Introduction

The use of wild relatives in crop improvement

Crop wild relatives provide a rich source of genetic diversity for breeding in cultivated crops. Some examples of valuable traits from wild relatives that have been transferred to crops include pest and disease resistance, yield increase, abiotic stress tolerance, quality traits, and cytoplasmic male sterility (CMS) and male fertility restorers.

Based on a review study by Hajjar and Hodgkin (2007), more than 80% of crop improvements from wild relatives have incorporations of insect and disease resistance genes, such as resistance to potato late blight from *Solanum demissum* Lindl., stem and leaf rusts in wheat from *Agropyron elongatum* Host ex. P.Beauv and *Aegilops umbellulata* Zhuck., and grassy stunt virus in rice from *Oryza nivara*. Yield increase from the use of wild species is reported in many species including rice, chickpea and cassava, but mostly as a result of disease resistance provided by the wild genes. The available case study of a released cultivar using wild germplasm to specifically increase yield is rice cultivar NSICRc112 from the cross *O. sativa x O. longistaminata* (Hajjar and Hodgkin 2007). Examples of other traits from crop wild relatives include resistance to abiotic stresses, such as drought tolerance in barley from *Hordeum spontaneum* K. Koch, improved quality traits such as increased soluble solid content, fruit color, and adaptation to mechanical harvesting in tomato, and the uses of CMS in sunflower, rice and millet (Hajjar and Hodgkin 2007).

Due to the importance of wild relatives to improvement of cultivated crops, genomic libraries of introgression lines have been developed in several crops, such as tomato, rice, wheat, rye, melon, and lettuce (Ali et al., 2010; Falke et al., 2009; Jeuken & Lindhout, 2003; Perpiñá et al., 2016). Nonetheless, the most notable and successful stories are the development of several sets of ILs from tomato wild relatives, such as *S. lycopersicoides, S. pennellii, S. pimpinellifolium, S. habrochaites, S. hirsutum, S. peruvianum* and *S. sitiens*. (Bai & Lindhout, 2007; Canady et al., 2005; Eshed & Zamir, 1994; Yates et al., 2004). More than 40 disease resistance genes have been introgressed into cultivated tomatoes from 10 wild relatives with virtually all the disease resistance genes currently in commercial cultivars. Abiotic stress genes and quality trait genes from two wild relatives for 55 total contributed traits have also been introgressed into cultivated tomatoes (Hajjar and Hodgkin 2007).

Cucumber

Taxonomy, origin and production

The Cucurbitaceae family hosts about 1,000 species in 96 genera including *Cucumis* (cucumber and melon), *Cucurbita* (squash and pumpkin), *Citrullus* (watermelon), and *Luffa* (loofah). The genus *Cucumis* contains 66 species; most of them are monoecious but a few, such as *Cucumis hirsutus*, are dioecious (Renner and Schaefer 2016). Cucumber is native to southern Asia (de Candolle, 1959; Sebastian et al., 2010). Four botanical varieties of *C. sativus* have been identified which include the cultivated cucumber *C. sativus* var. *sativus*, the wild cucumber *C. sativus* var. *hardwickii,* the semi-wild Xishuangbanna cucumber *C. sativus* var. *xishuangbannesis,* and the Sikkim cucumber *C. sativus* var. *sikkimensis* (Bo *et al.*, 2015; Renner & Pandey, 2013; Hooker, 1876).

Cucumber is one of the important vegetable crops cultivated worldwide and ranks among the top 20 vegetable crops. In the US, grown with $333 million production value in the United States. Total cucumber production in 2018 was over 700,000 metric tons, of which 284,493 tons was for the fresh market and 491,268 tons for processing (USDA as of March 2019). The main producers of cucumber are in Asia (87.9%, mostly in China) and Europe (7.4%), while the U.S. shares about 2.9% of market production share as of 2017 (FAOSTAT, 2019). There are several market types of cucumber production based on how cucumbers are grown and consumer preferences in different cultures around the world. Notable cucumber market types include North American slicers and processing types, North China fresh market type and cucumber greenhouse type grown in Europe (Naegele and Wehner 2017).

Genetic diversity in cucumber

Cultivated cucumbers within each market class have a very narrow genetic base with 3-20% polymorphism level revealed with molecular markers (Cavagnaro et al., 2010; Kang et al., 2011; Lv et al., 2012; Weng et al., 2010). Cucumber lines, and in many cases QTLs and candidate genes, have been identified for traits including increased yield, plant architecture, seedling traits, sex flower expression, fruit shape, and disease resistance (Naegele and Wehner 2017). Resistant cultivars to downy mildew (DM, caused by *Pseudoperonospora cubensis*), powdery mildew (PM, caused by *Podosphaera xanthii*), cucumber mosaic virus (CMV), zucchini yellow mosaic virus (ZYMV) and other disease are available in cucumber (Fukino et al. 2013; He et al. 2013; Wang et al. 2016). However, search for additional resistances from other gene pools is crucial to prevent possible resistance breakdown in the future. Thus, increasing genetic diversity of the cucumber primary gene pool is critical to aid in cucumber breeding.

C. hystrix as the wild source for cucumber improvement

The secondary gene pool for species in *Cucumis* genus, which are either cross incompatible (such as *C. africanus, C. anguria, C. dipsaceus,* and *C. metuliferus*) or sparingly cross compatible (such as its wild relative *C. hystrix*). The tertiary gene pool for cucumber are distantly related species from other genera or sub-genera (such as *C. melo* and *Cucurbita* L.) that do not hybridize with cucumber (Naegele and Wehner 2017). Early attempts of interspecific hybridization to transfer useful genes from wild relatives into cucumber or melon were not successful (reviewed by Chen & Adelberg 2000). For example, high level SRKN resistance was identified in *C. metuliferus* (Fassuliotis 1979; Nugent & Dukes 1997; Walters et al. 1993, 2006). However, no viable interspecific hybrids of cucumber or melon with this species have been obtained thus far (e.g., Deakin et al. 1971; Kroon et al. 1979; Kho et al. 1980; Norton & Barker 1980).

Cucumis hystrix is the sister species to cucumber and is by far the only known species that is cross-compatible with cucumber (Chen et al., 1997). *C. hystrix* possesses a number of valuable traits, such as abiotic stress to cold tolerance (field observation) and disease resistance. Several disease resistance screening tests indicated that *C. hystrix* is resistance to downy mildew (DM, caused by *Pseudoperonospora cubensis*), root-knot nematode (RKN, caused by *Meloidogyne incognita*), gummy stem blight (caused by *Didymella bryoniae*), cucumber mosaic virus (CMV), zucchini yellow mosaic virus (ZYMV), and watermelon mosaic virus (WMV) (Chen et al., 2004; Chen & Zhou, 2011; Jiang et al., 2010). Therefore, *C. hystrix* should serve as a potential source of secondary gene pool for cucumber improvement.

Cucumis hystrix

Taxonomy and origin

Cucumis hystrix belongs to the Cucurbitaceae family and *Cucumis* genus. It is a diploid species with 12 chromosomes similar to melon, *C. melo*, and several other wild species in the *Cucumis* genus, such as *C. debilis, C. anguria*, and the horned melon, *C. metuliferus*. *C. hystrix* is the sister species to cucumber and is the only wild species that is grouped into subgenus *Cucumis* with cucumber (Kirkbride Jr., 1993; Renner et al., 2007). Studies have suggested that *C. hystrix* is the closest relative to wild and cultivated cucumbers and diverged from the common ancestor about 4.6 million years ago. *C. hystrix* originated in Asia with a distribution range in India, China, Myanmar, and Thailand (partly overlapped with cucumber) (Ghebretinsae et al., 2007; Paris, 2016; Renner et al., 2007; Sebastian et al., 2010).

Morphology

Cucumis hystrix was originally recorded by Chakravarty as an herb (1952). *C. hystrix* is an annual climber that grows many branches (**Figure 1.1 A**) and has hairy stems (0.3-0.8 mm). Leaves are similar to cucumber though smaller in size (~7.5 cm x 5.5 cm). *C. hystrix* is monoecious with photoperiod-sensitive flowering time. Flower appearance is similar to cucumber in appearance but much smaller in size. A male flower is ~1-2 cm long (**Figure 1.1 B**) and a female flower is ~3-4 cm long. Internode length is about 4-6 cm. Seeds are tiny, about 1.4 mm wide and 2.56 mm long (**Figure 1.1 C**). Fruits are dark-green and oval shaped with black, spiny tubercles (warts) (**Figure 1.1 D**). *C. hystrix* fruits are mostly consisted of seeds and rind when cutting cross-section, but they are edible and a bit sour in taste. While *C. hystrix* possesses

a number of valuable traits, its fruit characteristics (wart and black spines) could serve as a linkage drag for cucumber improvement.

Research Objectives

This research project was undertaken with four main objectives. The first objective was to develop a draft genome assembly of *C. hystrix* using high throughput sequencing to obtain better coverage for highly accurate molecular analyses, such as genetic mapping and comparative study between *Cucumis* species (Chapter 2). The second objective was to construct a linkage map for *C. hystrix* from a SNP dataset generated by genotyping-by-sequencing (GBS) technology, to combine the SNP markers with a previously developed linkage map, and to anchor scaffolds from the draft genome assembly to generate an integrated genetic-physical map of *C. hystrix* genome. We then used the map and the karyotype for *C. hystrix* (Koo, *unpublished*) to refine the syntenic relationship between cucumber and *C. hystrix* chromosomes (Chapter 3). The third objective was to develop an introgression library of *C. hystrix* in the elite cucumber Gy14 and 9930 genetic backgrounds with marker-assisted backcrossing, using the linkage map developed in objective (2) to detect the presence of *C. hystrix* introgression segments (Chapter 4). The fourth and last objective was to characterize ILs molecularly and phenotypically for multiple traits with a focus on disease resistances, such as downy mildew, angular leaf spot, and root-knot nematode (Chapter 5). The ILs developed will be a valuable source for broadening the genetic base of cucumber, disease resistance breeding, discoveries and mapping of horticulturally important QTLs in cucumber.

Reference

Ali ML, Sanchez PL, Yu S, et al (2010) Chromosome Segment Substitution Lines: A Powerful Tool for the Introgression of Valuable Genes from *Oryza* Wild Species into Cultivated Rice (*O. sativa*). Rice 3:218–234. doi: 10.1007/s12284-010-9058-3

Bai Y, Lindhout P (2007) Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? Ann Bot 100:1085–94. doi: 10.1093/aob/mcm150

Bo K, Ma Z, Chen J, Weng Y (2015) Molecular mapping reveals structural rearrangements and quantitative trait loci underlying traits with local adaptation in semi-wild Xishuangbanna cucumber (*Cucumis sativus* L. var. *xishuangbannanesis* Qi et Yuan). Theor Appl Genet 128:25–39. doi: 10.1007/s00122-014-2410-z

Canady MA, Meglic V, Chetelat RT (2005) A library of *Solanum lycopersicoides* introgression lines in cultivated tomato. Genome 48:685–97. doi: 10.1139/g05-032

Cavagnaro PF, Senalik DA, Yang L, et al (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). BMC Genomics 11:1–18. doi: 10.1186/1471-2164-11-569

Chakravarty HL (1952) New finds of Indian Cucurbitaceae. J Bombay Nat Hist Soc 50:894–901

Chen J, Moriarty G, Jahn M (2004) Some disease resistance tests in *Cucumis hystrix* and its progenies from interspecific hybridization with cucumber. In: Progress in cucurbit genetics and breeding research. Proceedings of Cucurbitaceae 2004, the 8th EUCARPIA meeting on Cucurbit Genetics and Breeding. pp 189–196

Chen J, Staub JE, Tashiro Y, et al (1997) Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr. Euphytica 96:413–419. doi: 10.1023/A:1003017702385

Chen J, Zhou XH (2011) Wild Crop Relatives: Genomic and Breeding Resources (Vegetables). In: Chittaranjan K (ed) Wild Crop Relatives: Genomic and Breeding Resources Vegetables. Springer, pp 67–87

Eshed Y, Zamir D (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: A tool for fine mapping of genes. Euphytica 79:175–179

Falke KC, Miedaner T, Frisch M (2009) Selection strategies for the development of rye introgression libraries. Theor Appl Genet 119:595–603. doi: 10.1007/s00122-009-1069-3

FAOSTAT (2019). Available at: http://www.fao.org/faostat/en/#data/QC/visualize

Fukino N, Yoshioka Y, Sugiyama M, et al (2013) Identification and validation of powdery mildew (*Podosphaera xanthii*)-resistant loci in recombinant inbred lines of cucumber (*Cucumis sativus* L.). Mol Breed 32:267–277. doi: 10.1007/s11032-013-9867-3 Identification

Ghebretinsae AG, Thulin M, Barber JC (2007) Relationships of cucumbers and melons unraveled: Molecular Phylogenetics of *Cucumis* and related genera (Benincaseae, Cucurbitaceae). Am J Bot 94:1256–1266. doi: 10.3732/ajb.94.7.1256

Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica 156:1–13. doi: 10.1007/s10681-007-9363-0

He X, Li Y, Pandey S, et al (2013) QTL mapping of powdery mildew resistance in WI 2757 cucumber (*Cucumis sativus* L.). Theor Appl Genet 126:2149–2161

Hooker JD (1876) *Cucumis sativus* var. *Sikkimensis*. Botanical Magazine 6206

Jeuken MJW, Lindhout P (2003) Future perspectives of Backcross Inbred Lines for exploitation of wild germplasm: a case study on *Lactuca saligna* as a donor for quantitative resistance to lettuce downy mildew. Eucarpia Leafy Veg 69–74

Jiang B, Wu Z, Lou Q, et al (2010) Genetic diversity of Ty1-*copia* retrotransposons in a wild species of *Cucumis* (*C. hystrix*). Sci Hortic (Amsterdam) 127:46–53. doi: 10.1016/j.scienta.2010.09.013

Kang H, Weng Y, Yang Y, et al (2011) Fine genetic mapping localizes cucumber scab resistance gene *Ccu* into an *R* gene cluster. Theor Appl Genet 122:795–803. doi: 10.1007/s00122-010-1487-2

Lv J, Qi J, Shi Q, et al (2012) Genetic Diversity and Population Structure of Cucumber (*Cucumis sativus* L.). PLoS One 7: e46919. doi: 10.1371/journal.pone.0046919

Naegele RP, Wehner TC (2017) Genetic Resources of Cucumber. In: Grumet R, Katzir N, Garcia-Mas J (eds) Genetics and Genomics of Cucurbitaceae. Plant Genetics and Genomics: Crops and Models, vol 20. Springer International Publishing, pp 61–86

Paris HS (2016) Overview of the origins and history of the five major cucurbit crops: issues for ancient DNA analysis of archaeological specimens. Veg Hist Archaeobot 25:405–414. doi: 10.1007/s00334-016-0555-1

Perpiñá G, Esteras C, Gibon Y, et al (2016) A new genomic library of melon introgression lines in a cantaloupe genetic background for dissecting desirable agronomical traits. BMC Plant Biol 16:1–21. doi: 10.1186/s12870-016-0842-0

Renner SS, Pandey AK (2013) The Cucurbitaceae of India: Accepted names, synonyms, geographic distribution, and information on images and DNA sequences. PhytoKeys 53–118. doi: 10.3897/phytokeys.20.3948

Renner SS, Schaefer H (2016) Phylogeny and Evolution of the Cucurbitaceae. In: Grumet R, Katzir N, Garcia-Mas J (eds) Genetics and Genomics of Cucurbitaceae. Plant Genetics and Genomics: Crops and Models, vol 20. pp 13–23

Renner SS, Schaefer H, Kocyan A (2007) Phylogenetics of *Cucumis* (Cucurbitaceae): Cucumber (*C. sativus*) belongs in an Asian/Australian clade far from melon (*C. melo*). BMC Evol Biol 7:1–11. doi: 10.1186/1471-2148-7-58

Sebastian P, Schaefer H, Telford IRH, Renner SS (2010) Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. Proc Natl Acad Sci 107:14269–14273. doi: 10.1073/pnas.1005338107

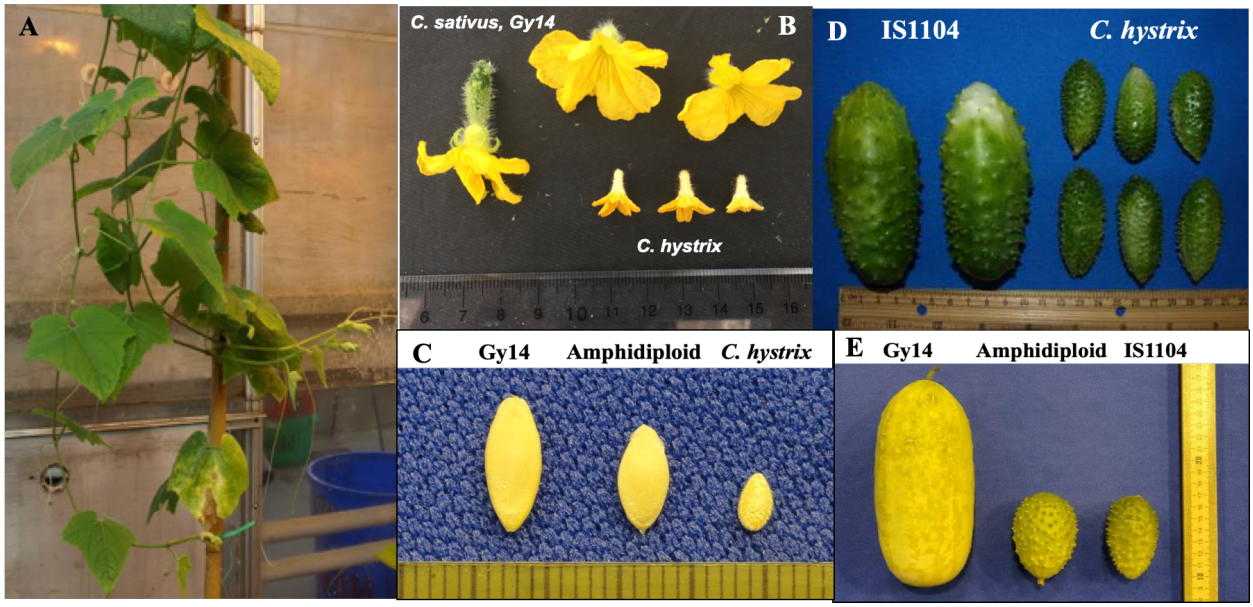
United States Department of Agriculture National Agricultural Statistics Service (2019) Vegetables 2018 Summary. Retreived from https://downloads.usda.library.cornell.edu/usda-esmis/files/02870v86p/gm80j322z/5138jn50j/vegean19.pdf

Wang Y, VandenLangenberg K, Wehner TC, et al (2016) QTL mapping for downy mildew resistance in cucumber inbred line WI7120 (PI 330628). Theor Appl Genet 129:1493–1505. doi: 10.1007/s00122-016-2719-x

Weng Y, Johnson S, Staub JE, Huang S (2010) An extended intervarietal microsatellite linkage map of cucumber, *Cucumis sativus* L. HortScience 45:882–886. doi: http://hdl.handle.net/10113/43470

Yates HE, Frary A, Doganlar S, et al (2004) Comparative fine mapping of fruit quality QTLs on chromosome 4 introgressions derived from two wild tomato species. Euphytica 135:283–296. doi: 10.1023/B:EUPH.0000013314.04488.87

Figure 1.1 *Cucumis hystrix* plant morphology. A) Plants and leaves. B) *C. hystrix* male flowers compared to male and female flowers of cucumber Gy14 accession. C) Seeds of *C. hystrix* compared to cucumber Gy14 accession and an amphidiploid from a cross between *C. hystrix* x Gy14. D) *C. hystrix* fruit size compared to an interspecific F_1 hybrid (IS1104) from a cross between *C. hystrix* x Gy14. E) Fruit size of IS1104 compared to cucumber Gy14 accession and an amphidiploid.



Chapter 2: *De novo* assembly of *C. hystrix* draft genome using next-generation sequencing

Abstract

Cucumis hystrix (2*n* = 2*x* = 24, HH) is the only known species in the genus *Cucumis* that is cross-compatible with cucumber (*C. sativus*, 2*n* = 2*x* = 14, CC) and has a great potential for cucumber improvement. To facilitate introgression of *C. hystrix* chromatins into cucumber genetic background, we sequenced two accessions, TH1 and CN1, of the *C. hystrix* genome with Illumina Hi-Seq 2000 and Roche/454 platforms and conducted *de novo* draft genome assembly. With the Newbler (v2.8) software package, 225 out of 378 million reads (60%) were assembled for TH1. The estimated genome size was 447.0 Mb. The resulting assembly contained 16,741 scaffolds (~78× coverage) for a total of 226.0 Mb, representing ~50.6% of the estimated 447 Mb *C. hystrix* genome. The largest scaffold being 342 kb and N50 scaffold size of 23.3 kb. For the CN1 accession, 72.4 out of 117 million reads (61.9%) were assembled with 24× coverage and an estimated genome size of 478.8 Mb. The assembly contains 88,371 contigs with N50 contig size of 4.0 kb. Scaffolding of the CN1 genome was not obtained due to quality of Roche/454 reads. The TH1 assembly was assessed for completeness with BUSCO to be 88.2% complete, which was used as the reference genome in developing a linkage map for *C. hystrix.* Characterization of repetitive DNA elements revealed that *C. hystrix* had similar proportion of 45S rDNA, 5S rDNA, and overall repeat elements as compared to cucumber.

Introduction

C. hystrix is a diploid species with a chromosome number *2n = 2x* = 24, similar to melon, *C. melo,* and several other wild species in the *Cucumis* genus, such as *C. debilis, C. anguria,* and horned melon, *C. metuliferus*. *C. hystrix* possesses a number of valuable traits, such as resistance to downy mildew (DM), root-knot nematode (RKN), gummy stem blight, cucumber mosaic virus (CMV), and zucchini yellow mosaic virus (ZYMV) (Chen et al. 2004; Jiang et al. 2010; Chen and Zhou 2011). Cucumber and *C. hystrix* were originated in Asia (Ghebretinsae et al., 2007; Paris, 2016; Renner et al., 2007; Sebastian et al., 2010). *C. hystrix* is the sister species to cucumber and is the only known species that is cross-compatible with cucumber (Chen et al., 1997). Thus, *C. hystrix* could be a potential source of secondary gene pool for cucumber improvement.

However, genomic resources available in *C. hystrix* are very limited. A previous draft genome assembly and a genetic map for *C. hystrix* were developed by Yang et al. (2014), in which hybrid assembly of *C. hystrix* accessions WI7001 (TH1) and WI7002 (CN1) yielded 11,649 scaffolds with a total of 209 Mb sequences. The largest scaffold was 560,870 bp with an N50 scaffold size of 50,831 bp. A genetic map for *C. hystrix* comprising of 416 simple sequence repeat (SSR) marker loci which anchored 197 scaffolds and 109 contigs (438 kb total or about 0.21% of the assembly), indicating that the assembly was fragmented. A small number of markers and insufficient coverage of the genome assembly (33.5x Illumina and 3.6X Roche/454) resulted in an extremely low level of scaffold anchoring to the genetic map (Yang et al. 2014).

The advancement of next generation sequencing (NGS) technology allows most organisms of interest to be sequenced quickly with high accuracy. At present, a genome could be sequenced at less than $1,000 per gigabase (Gb) (Goodwin et al., 2016). Current and newly developed sequencing platforms, such as Single Molecule, Real-Time (SMRT) sequencing and NanoPore sequencing, have the ability to produce millions of DNA sequence reads from a single run in a few hours. NGS provides an inexpensive, genome-wide tool to approach biological problems ranging from assembling a genome *de novo,* mutation mapping to noncoding RNA discovery and other functional genomic studies. In this study, we aimed to develop a new draft genome assembly of *C. hystrix* using high throughput sequencing to get better depth coverage for subsequent genetic mapping and comparative study between *Cucumis* species.

Materials and Methods

Plant materials and whole genome sequencing

Two *C. hystrix* lines TH1 and CN1 (accessions WI7001 and WI7002, respectively) were used for whole-genome sequencing. WI7001 was originally collected from Thailand and WI7002 was collected from Southwestern China. Both accessions were maintained through cuttings and grown in the Walnut Street Greenhouse, University of Wisconsin-Madison. Genomic DNA samples were extracted using CTAB method and purified using phenol/chloroform/isoamyl alcohol following the protocol presented in Murray and Thompson (1980). Both accessions were whole-genome sequenced with Roche/454 platform and Illumina Hi-Seq 2000 platform with 2x100 bp, 500-bp fragment insert at the Biotechnology Center, University of Wisconsin, Madison in 2011. For the 454 GS-FLX Titanium pyrosequencing, 8-kb paired-end libraries were used. WI7001 or TH1 accession were sequenced for two more lanes with paired-end Illumina

Hi-Seq 2000 platform with 2x100 bp, 500-bp fragment at the Beijing Genome Institute, China in 2014 to improve accuracy and increase coverage of the draft genome assembly.

Sequence read quality control

For Illumina Hi-Seq 2000 sequences, raw reads with quality scores less than 28 and length shorter than 41 bp were removed to obtain high quality reads using FastQC (Andrews, 2010) and bb.fastqfilter programs (Senalik, 2010).

For Roche/454 sequences, reads were converted from .sff format to standard FASTQ format with the bb.sff2fastq program (Das, 2010) and assessed for quality scores using FastQC program. The reads were trimmed by FASTX-Toolkit program version 0.0.14 with the fastx_trimmer function (http://hannonlab .cshl.edu/fastx_toolkit/) at both ends to remove over represented k-mers as assessed by FastQC program. Low quality reads were removed with Trimmomatric program version 0.32 (Bolger et al. 2014) if the average quality score fell below a threshold of 25 for every 15-nucleotide window. The quality score was lowered to 25 in this case because most reads were trimmed to 100-200 bp when setting the quality score at 28, which was too short and too much information was lost.

Draft genome assembly of C. hystrix

The *C. hystrix* genome was assembled into two versions – one for each accession: TH1 and CN1 assemblies. In TH1 assembly, Illumina (3 lanes) and Roche/454 raw reads were used as described above. One lane Illumina and Roche/454 raw reads were used for CN1 assembly. The draft genome was assembled *de novo* using the Newbler version 2.8 software package (2013). Minimum overlap length was set at 45 bp (default: 40 bp) and minimum overlap identity of reads

was set at 95% (default: 90%). The rest of the assembly parameters were set to defaults.

Contamination removal from the draft genome assembly

We discovered possible non-cucumber sequence contamination in the draft genome assembly when calling SNPs in the tassel pipeline (Chapter 3) using the assembly as the reference genome. To detect and remove contamination, two largest scaffolds (467 kb and 405 kb, respectively) in the *C. hystrix* TH1 assembly that failed to call any SNPs were run against the cucumber protein and EST database (Clark et al., 2016) using the MAKER annotation pipeline with EST2gene option (Campbell et al., 2014). Regions that match proteins in the database were extracted and identified with BLASTn (Altschul et al., 1990). Once the sources of the contamination were identified, the entire draft assembly was BLASTed against them. The draft genome assembly was further blasted against the *C. hystrix* plastid genome (source: https://www.ncbi.nlm.nih.gov/nuccore/NC_023544.1/), *C. sativus* mitochondrion genome chromosome 1 (source: https://www.ncbi.nlm.nih.gov/nuccore/NC_016005.1), chromosome 2 (source: https://www.ncbi.nlm.nih.gov/nuccore/NC_016004.1), and chromosome 3 (source: https://www.ncbi.nlm.nih.gov/nuccore/NC_016006.1). Scaffolds with more than 50% match in length to the contamination sources were removed. The remaining scaffolds were considered clean enough for further analysis.

Assessing genome assembly and annotation completeness with single-copy orthologs

The cleaned draft genome assembly (version 5.1, *C. hystrix* TH1 genome) was further assessed for quality and annotation completeness with a software called Benchmarking Universal Single-Copy Orthologs or BUSCO version 2. BUSCO compares an assembly to a set of genes known to

exist across each kingdom or phylum, such as plants, nematodes or protists as an evolutionary measure of genome completeness. If the assembly contains most of the genes existing in that kingdom or phylum, then the assembly is more complete (Simão et al., 2015).

Characterization of repetitive DNA elements in C. hystrix

Illumina Hi-Seq 2000 sequencing reads were used as input to identify and characterize repetitive elements in *C. hystrix* genome. The analysis was performed using RepeatExplorer software on a Galaxy-based web server (http://repeatexplorer.org/) (Novák *et al.*, 2013) using Viridiplantae repeat database and default parameters.

Results

Sequence read quality control

For Illumina reads (2011), most of the reads passed minimum quality score and minimum length thresholds. For TH1 Illumina reads, 979,250 sequences or 0.72% of 135,870,754 sequences did not pass the thresholds and were rejected. For CN1 Illumina reads, 1,004,464 sequences or 0.78% of 117,104,464 sequences did not pass the thresholds and were rejected. For Illumina reads (2014), all reads from both TH1 and CN1 accessions passed the quality thresholds.

For Roche/454 paired-end reads, FastQC analysis revealed that the reads at both ends had really low quality and were trimmed (**Figure 2.1A**). With trimmomatic software, TH1 reads were trimmed to ~410 bp in length for quality score of 25 and sliding window of 15 bp. Out of 430,225 reads, 336,038 (78.11%) read pairs, 34,921 (8.12%) forward reads, and 53,345 (12.40%) reverse reads remained. For CN1, reads were trimmed to ~450 bp in length for quality score of

25 and sliding window of 15 bp. Out of 514,404 reads, 338,265 (65.75%) read pairs, 73,454 (14.28%) forward reads, and 83,555 (16.24%) reverse reads remained. FastQC analysis after trimming showed that the reads had good quality for use in the *de novo* assembly (**Figure 2.1B**).

Draft genome assemblies for C. hystrix

Using the Newbler (v2.8) software package, 225 out of 378 million reads (60%, Illumina and Roche/454 reads combined) were assembled for TH1. The software gave an estimated genome size of 447.0 Mbp from the resulting assembly. The assembly contained 16,865 scaffolds (~78× coverage) for a total of 229.6 Mb, representing about 51% of the estimated genome size for *C. hystrix* genome. The largest scaffold was 467 kb and N50 scaffold size was 23.8 kb.

For the CN1 accession, 72.4 out of 117 million reads (61.9%) were assembled with 24× coverage and an estimated genome size of 478.8 Mbp. The assembly contained 88,371 contigs with N50 contig size of 4.0 kb. Scaffolds for CN1 could not be obtained because too much paired-end information from Roche/454 reads were lost after quality trimming. Only 65.75% of the Roche/454 read input for the assembly had both reads in a pair. Thus, the *C. hystrix* CN1 assembly was not considered in subsequent work (**Table 2.1**).

Using the MAKER annotation pipeline, it was found that the two largest scaffolds from *C. hystrix* TH1 assembly were sequences from the gram-negative bacterial species *Legionella pneumophila subsp. pneumophila str. Birmingham 1 (D-7470)*. The DNA sample sent for sequencing could have been contaminated from soil or water in the greenhouse. Scaffolds with more than 50% reads from bacterial DNA or *C. hystrix* chloroplast DNA or *C. sativus*

mitochondrion DNA were removed (124 scaffolds or 3.6 Mbp total).

The cleaned assembly version (TH1 accession) contained 16,741 scaffolds for a total of 226.0 Mb, representing 50.6% of the estimated genome size. The largest scaffold was 342 kb, N50 scaffold 23.3 kb, and an estimated genome size of 447.0 Mb (**Table 2.1**). One hundred and sixty-six scaffolds (0.99%) were larger than 100 kb, 559 scaffolds (3.34%) were between 50-100 kb, 5,610 scaffolds (33.51%) were between 10-50 kb, and 10,406 scaffolds (62.16%) were smaller than 10 kb in size. This version of cleaned assembly was for quality assessment by BUSCO.

Assessing genome assembly and annotation completeness

The draft genome assembly (TH1) was assessed by BUSCO software. Against 956 BUSCO ortholog groups, the assembly contains 843 complete BUSCOs (88.2%); 819 of which are complete and single-copy BUSCOs (85.7%) and 24 of which are complete and duplicated BUSCOs (2.5%). Fifty-four ortholog groups are fragmented BUSCOs (5.6%) and 59 ortholog groups are missing from this assembly (6.2%).

Characterization of repetitive DNA elements in C. hystrix

C. hystrix possessed about 23.85% of repetitive DNA elements in its genome (**Table 2.2**). The majority of the repetitive elements were unclassified (17.24%). The proportion of 45S rDNA and 5S rDNA was 3.01% and 0.23%, respectively. The genome proportion of mobile element class I, which consisted of LTR Ty1-*copia* and LTR Ty3-*Gypsy* and organellar genome, which consisted of plastid and mitochondria, were at similar level (0.05% and 0.04%, respectively). The proportion of repetitive DNA elements in cucumber genome was derived from Han et al. (2008)

and Huang et al. (2009) and from running cucumber Gy14 Illumina reads with RepeatExplorer software.

Discussion

Genome assembly and annotation completeness

A previous draft genome assembly for *C. hystrix* developed by Yang et al. (2014) contained 11,649 scaffolds and totaled 209 Mbp in length with 33.5x Illumina and 3.6X Roche/454 coverage. The largest scaffold was 560,870 bp with an N50 scaffold size of 50,831 bp. Overall, the new *C. hystrix* assembly in this study presented an improvement. It had higher coverage depth (78x) and longer assembly length (20 Mb longer). The largest scaffold (467 kb) and N50 scaffold size (23.8 kb) were slightly lower than the previous assembly. The reason for shorter scaffold length was that Roche/454 reads were not quality trimmed in the previous assembly. The reads were trimmed if quality score fell below 25 for every 15-nt sliding window, resulting in shorter read length and some paired-end read information lost as a trade-off for better quality read input in the assembly process.

The genome completeness assessed by BUSCO was 88.2% complete, which was lower than that of the *C. sativus* Gy14 version 2 assembly (94.67%) (Xu et al., personal communication) and the *Cucumis melo* version 3.6.1 (95.08%) (Ruggieri et al. 2018). For comparison, the completeness of the genome assemblies for *Arabidopsis thaliana*, *Cicer arietinum* (chickpea), *Nelumbo nucifera* (Indian lotus), *Lolium perenne* (ryegrass) and *Oryza sativa* ssp. *japonica* (rice) were higher than 95% (Veeckman, E. et al., 2016).

The Newbler software also produced an "edge" statistic. Edge represents reads that exit a contig or scaffold and enter another one. Contigs or scaffolds with no edges indicate that there are not enough reads to bridge reads or that the coverage is too low on those contigs or scaffolds (Nederbragt, 2010). For *C. hystrix* TH1 assembly, 31.40% of all scaffolds had no edges, 48.30% had one edge (reads that connect contigs or scaffolds at one end only), and 9.20% had both edges. This indicated that the assembly was quite fragmented and probably explains why only 26.62% of the scaffolds were anchored to the genetic map (see Chapter 3). Despite these shortcomings, we felt that the clean assembly was of sufficient quality for further downstream analysis, specifically for SNP calling from genotyping-by-sequencing (GBS) to construct a *C. hystrix* genetic map in Chapter 3 and to identify *C. hystrix* introgressions in Chapter 4.

Repetitive DNA elements in C. hystrix

The proportion of 45S rDNA, 5S rDNA and total repetitive DNA elements in the *C. hystrix* genome was similar to that of cucumber (~24% total repetitive DNA in cucumber) (Han et al. 2008; Huang et al. 2009) but lower than the estimate in melon (34% of the genome) (Ruggieri et al. 2018). Class I transposons was the most abundant transposable elements in melon (33.2% of the melon genome), while represented only 0.05% of the *C. hystrix* genome. This result suggested that *C. hystrix* genome is closer to cucumber than to melon despite similar estimated genome size to melon (447 Mb for *C. hystrix* and 450-500 Mb for *C. melo*) (Arumuganathan and Earle 1991), similar to phylogenetic studies in *Cucumis* (Paris, 2016; Renner et al., 2007; Sebastian et al., 2010).

Moreover, Han et al (2008) also demonstrated that the distribution of Type I/II and Type IV repeats was similar between *C. hystrix* and cucumber. However, the estimate satellite proportion of *C. hystrix* (11.07%) was much lower than that of cucumber (19.96%). The genome proportions of LTR Ty1-*copia* and LTR Ty3-*gypsy* were vastly different as well (**Table 2.2**). Based on Gy14 Illumina sequence reads, LTR Ty1-*copia* represented about 10.70% of the cucumber genome. This is comparable to the estimated in *C. hystrix* at 8.5% in previous study (Jiang et al. 2010). The estimated number of LTR Ty1-*copia* in Jiang et al's study was calculated from Southern hybridization, while the estimated number of LTR Ty1-*copia* and other repetitive elements in cucumber was from aligning reads to the repeat database. Larger genome in eukaryotes usually accumulate more repetitive elements (Bennetzen *et al.*, 2005; Feschotte *et al.*, 2002; Kidwell, 2002). Since the estimated genome size of *C. hystrix* is larger than that of cucumber, by definition, *C. hystrix* should have accumulated more repetitive elements in its genome. Low estimate of repetitive elements in *C. hystrix* suggested that *C. hystrix* Illumina sequence reads might align poorly to the repeat database.

Acknowledgement

I thank Kristin Haider for technical help. I also thank Dr. Douglas Senalik for providing a server space for assembling a genome and other recommendations regarding bioinformatic work.

References

454 Life Sciences Corp/Roche Company (2013). GS De Novo Assembler version 2.8. Brandford, CT

Andrews, S (2010). FastQC software. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Altschul SF, Gish W, Miller W, et al (1990) Basic Local Alignment Search Tool. J Mol Biol 215:403–410. doi: 10.1016/S0022-2836(05)80360-2

Arumuganathan K, Earle ED (1991) Nuclear DNA Content of Some Important Plant Species. Plant Mol Biol Report 9:208–218. doi: 10.1007/BF02672016

Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of Recent Genome Size Variation in Flowering Plants. Ann Bot 95:127–132. doi: 10.1093/aob/mci008

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. doi: 10.1093/bioinformatics/btu170

Campbell MS, Holt C, Moore B, Yandell M (2014) Genome Annotation and Curation Using MAKER and MAKER-P. In: Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp 4.11.1-4.11.39

Chen J, Moriarty G, Jahn M (2004) Some disease resistance tests in *Cucumis hystrix* and its progenies from interspecific hybridization with cucumber. In: Proceedings of Cucurbitaceae 2004, the 8th EUCARPIA Meeting on Cucurbit Genetics and Breeding. pp 189–196

Chen J, Staub JE, Tashiro Y, et al (1997) Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr. Euphytica 96:413–419. doi: 10.1023/A:1003017702385

Chen J, Zhou XH (2011) Wild Crop Relatives: Genomic and Breeding Resources (Vegetables). In: Chittaranjan K (ed) Wild Crop Relatives: Genomic and Breeding Resources Vegetables. Springer, pp 67–87

Clark K, Karsch-Mizrachi I, Lipman DJ, et al (2016) GenBank. Nucleic Acids Res 44:D67-72. doi: 10.1093/nar/gkv1276

Das, Indraniel (2010). Sff2fasq software. https://github.com/indraniel/sff2fastq

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: Where genetics meets genomics. Nat Rev Genet 3:329–341. doi: 10.1038/nrg793

Ghebretinsae AG, Thulin M, Barber JC (2007) Relationships of Cucumbers and Melons Unraveled: Molecular Phylogenetics of *Cucumis* and Related Genera (Benincaseae, Cucurbitaceae). Am J Bot 94:1256–1266. doi: 10.3732/ajb.94.7.1256

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17:333–351. doi: 10.1038/nrg.2016.49

Han YH, Zhang ZH, Liu JH, et al (2008) Distribution of the tandem repeat sequences and karyotyping in cucumber (*Cucumis sativus* L.) by fluorescence in situ hybridization. Cytogenet Genome Res 122:80–8. doi: 10.1159/000151320

Huang S, Li R, Zhang Z, et al (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41:1275–81. doi: 10.1038/ng.475

Jiang B, Wu Z, Lou Q, et al (2010) Genetic diversity of Ty1-copia retrotransposons in a wild species of *Cucumis* (*C. hystrix*). Sci Hortic (Amsterdam) 127:46–53. doi: 10.1016/j.scienta.2010.09.013

Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. Genetica 115:49–63. doi: 10.1023/A:1016072014259

Murray MG, Thompson WF (1980) Nucleic Acids Research Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4326. doi: doi:10.1093/nar/8.19.4321

Nederbragt, Lex (2010). How Newbler Works. https://contig.wordpress.com/2011/07/12/what-is-new-in-newbler-2-6/

Novák P, Neumann P, Pech J, et al (2013) RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–793. doi: 10.1093/bioinformatics/btt054

Paris HS (2016) Overview of the origins and history of the five major cucurbit crops: issues for ancient DNA analysis of archaeological specimens. Veg Hist Archaeobot 25:405–414. doi: 10.1007/s00334-016-0555-1

Renner SS, Schaefer H, Kocyan A (2007) Phylogenetics of *Cucumis* (Cucurbitaceae): Cucumber (*C. sativus*) belongs in an Asian/Australian clade far from melon (*C. melo*). BMC Evol Biol 7:1–11. doi: 10.1186/1471-2148-7-58

Ruggieri V, Alexiou KG, Morata J, et al (2018) An improved assembly and annotation of the melon (*Cucumis melo* L.) reference genome. Sci Rep 8:8088. doi: 10.1038/s41598-018-26416-2

Sebastian P, Schaefer H, Telford IRH, Renner SS (2010) Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister

species of melon is from Australia. PNAS 107:14269–14273. doi: 10.1073/pnas.1005338107

Senalik, D. (2010) bb.fastqfilter software. Internal use. University of Wisconsin-Madison.

Simão FA, Waterhouse RM, Ioannidis P, et al (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. doi: 10.1093/bioinformatics/btv351

Veeckman E, Ruttink T, Vandepoele K (2016) Are We There Yet? Reliably Estimating the Completeness of Plant Genome Sequences. Plant Cell 28:1759–1768. doi: 10.1105/tpc.16.00349

Yang L, Koo DH, Li D, et al (2014) Next-generation sequencing, FISH mapping and synteny-based modeling reveal mechanisms of decreasing dysploidy in *Cucumis*. Plant J 77:16–30. doi: 10.1111/tpj.12355

Table 2.1 Summary statistics of *C. hystrix* genome assemblies of TH1 and CN1 accessions

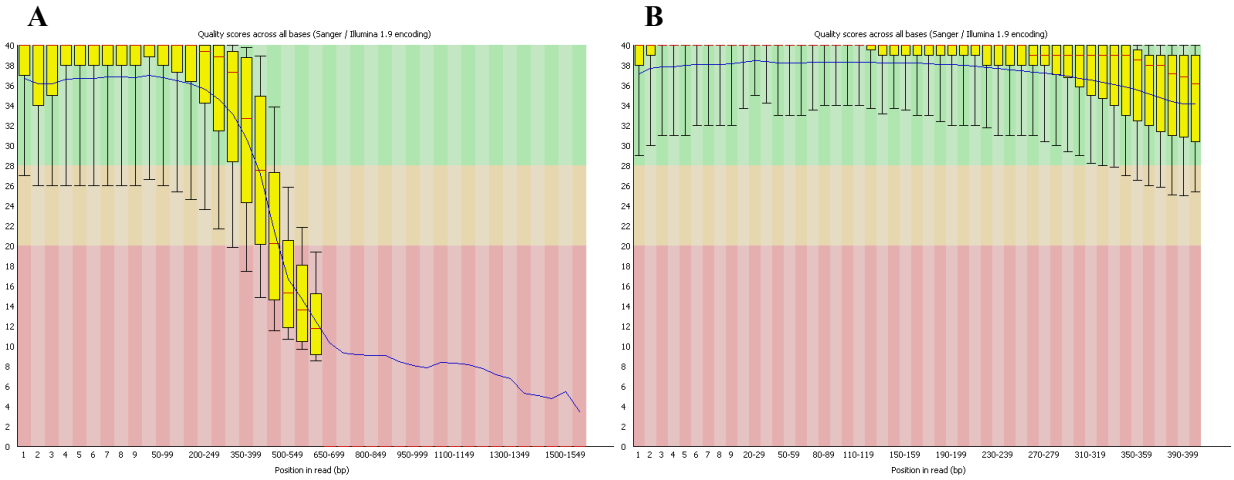
Assembly	**TH1**	**TH1 (clean)**	**CN1**
Initial number of reads	378,580,448	378,580,448	117,258,836
Percent of reads assembled (%)	59.56	59.56	61.90
Number of scaffolds	16,865	16,741	-
Average scaffold size (kb)	13.6	13.5	-
Maximum scaffold size (kb)	467.0	342.0	-
N50 scaffold (kb)	23.8	23.3	-
Number of large contigs > 2kb	54,085	-	88,371
Average contig size (kb)	4.0	-	2.6
Maximum contig size (kb)	211.9	-	36.4
N50 contig (kb)	7.0	-	4.0
Coverage (depth)	78	-	24
Total bases in assembly (bp)	229,643,897	225,955,978	228,043,770
Est. genome size (Mb)	447.0	447.0	478.8

Table 2.2 Characterization of repetitive DNA elements in *C. hystrix* TH1 accession from Illumina Hi-Seq 2000 reads

Repeat Type	**Genome proportion (%)**	**N_reads**	**Repeat proportion (%)**	**Estimates in cucumber**
45S rDNA	3.01	4,140	10.16	3.30 *
5S rDNA	0.23	321	0.79	0.13 *
Satellite	3.28	4,509	11.07	19.96 *
Mobile element class I	0.05			
LTR Ty1_Copia	0.03	46	0.11	10.70 †
LTR Ty3_Gypsy	1.62	2,223	5.46	7.50 †
Organelle	0.04			
Plastid	3.06	4,215	10.34	
Mitochondria	1.14	1,566	3.84	
Unclassified repeats	17.24	23,730	58.23	
Total	**23.85** †	**40,750**		

Note:
* Cucumber data from Han et al., 2008
† Estimated from Gy14 Illumina reads by RepeatExplorer software. Total repetitive DNA elements in cucumber as measured by RepeatExplorer was equal to 24.01%.

Figure 2.1 Roche/454 sequence read quality of *C. hystrix* TH1 assessed by FastQC (A) before and (B) after trimming with FASTX-Toolkit and Trimmomatric programs.



Chapter 3: Linkage map construction for *C. hystrix* and refinement of syntenic relationships between *C. hystrix* and cucumber chromosomes

Abstract

Cucumis hystrix (2*n* = 2*x* = 24, HH) is the only known species in the genus *Cucumis* that is sexually compatible with cucumber (*C. sativus*, 2*n* = 2*x* = 14, CC) and has a great potential for cucumber improvement. Resources in this species to aid in cucumber improvement are extremely limited. Here we reported a new SNP-based map and a consensus genetic map for *C. hystrix.* Through genotyping-by-sequencing (GBS), a linkage map for *C. hystrix* was constructed from 1,692 SNP loci, which was then integrated with a previously developed genetic map with 410 simple sequence repeat (SSR) markers. The resulting consensus map consisted of 12 linkage groups spanning 1119 cM, which anchored 1,321 scaffolds accounting for 60.2 Mb or ~26.62% of *C. hystrix* draft genome assembly. A karyotype for the *C. hystrix* genome was developed with molecular cytogenetic landmarks. These new genomic resources allowed refinement of the syntenic relationships among *C. hystrix*, cucumber and melon (*C. melo*, 2*n* = 2*x* = 24) chromosomes. An improved model was proposed to elucidate the evolutionary history in which the seven cucumber chromosomes were evolved from an *n*=12 ancestor through dysploid chromosome reduction that involved in fifty-nine syntenic blocks with four translocations, four chromosomal fusions, eighteen pericentric inversions and thirty-five paracentric inversions.

Introduction

Cucumis hystrix (*2n = 2x* = 24) is the closest sister species to cucumber and the only known species that is sexually compatible with cucumber (*2n = 2x* = 14) (Chen et al., 1997). Preliminary studies have shown that *C. hystrix* is resistant to downy mildew (DM), root-knot nematode (RKN), gummy stem blight, cucumber mosaic virus (CMV), and zucchini yellow mosaic virus (ZYMV) (Chen et al., 2004; Chen & Zhou, 2011; Jiang et al., 2010). Thus this species provides a great potential as a secondary gene pool for cucumber improvement.

However, exploiting the wild relative for cucumber improvement could prove difficult. *C. hystrix* and cucumber have different base chromosome number. Genetic resources available in *C. hystrix* for molecular studies are limited. Currently, only one *C. hystrix* genetic map was reported with 416 simple sequence repeat (SSR) marker loci, 215 of which were derived from *C. hystrix* (Yang et al., 2014). The map was only able to anchor 0.21% of the reported *C. hystrix* genome assembly. A small number of markers and insufficient coverage depth of the genome assembly (33.5x Illumina and 3.6X Roche/454) resulted in an extremely low level of scaffold anchoring to the genetic map. Other resources about *C. hystrix* were cytogenetic studies among *Cucumis* species, which revealed the genetic diversity of Ty1-*copia* retrotransposons (Jiang et al. 2010) and the distribution of several kinds of tandem repeats, including 45S rDNA, 5S rDNA, and centromeric repeats on *C. hystrix* chromosomes (Han et al., 2008; Li et al., 2016; Yang et al., 2017; Zhang et al., 2015).

The use of next-generation high throughput sequencing allows researchers to characterize the broad genomic structure of any species at a very low cost. The genotyping-by-sequencing (GBS)

technology can generate large number of markers for genomic studies with or without a reference genome, such as high-density genetic mapping, QTL detection, and genome-wide association studies (GWAS) (Elshire et al. 2011). Single-nucleotide polymorphisms (SNPs) allow for the construction of high-density linkage maps and could conveniently replace other types of markers, such as SSR markers for both major and minor crops simply because they can be generated efficiently and quickly using next-generation sequencing methods. In this study, we aimed to construct a high-density linkage map for *C. hystrix* with GBS to combine the SNP markers with a previously developed linkage map, and to anchor scaffolds from the new draft genome assembly to generate an integrated genetic-physical map of *C. hystrix* genome. We then used this map and the karyotype for *C. hystrix* (Koo, *unpublished*) to refine the syntenic relationship between cucumber and *C. hystrix* chromosomes.

Materials and Methods

Plant materials and population development

The F_2 population developed by Dr. Luming Yang was used for linkage map construction for this research (Yang et al., 2014). Two *C. hystrix* lines TH1 and CN1 (accessions WI7001 and WI7002, respectively) were used as the parental lines to develop an F_2 population in this study. The wild accession WI7001 was originally collected from Thailand and WI7002 was collected from Southwestern China. An F_1 plant from a cross between TH1 x CN1 was self-pollinated to produce 130 F_2 progenies.

DNA isolation and genotyping-by-sequencing

The first or second true leaf of *C. hystrix* F_2 seedlings were collected and vacuum-dried in a

freeze-dryer. The samples then were grounded into fine powder with a high-throughput homogenizer (OPS Diagnostics, Lebanon, NJ). Genomic DNA was isolated using CTAB method and purified using phenol/chloroform/isoamyl alcohol (Murray and Thompson 1980). The DNA was quantified using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Tech, Wilmington, DE) and diluted to a concentration of 50 ng/μL. Since the *C. hystrix* genetic map in Yang et al. (2014) was constructed from 91 F_2 individuals, 39 extra F_2 samples were genotyped using the same set of 412 SSR markers (Yang et al., 2014) to develop an SSR-based genetic map with 130 F_2 individuals. PCR amplification with SSR markers and gel electrophoreses of PCR products were conducted following Li et al (2011). Genotyping results with SSR markers were later combined with previous genotyping data on 91 F_2 samples (Yang et al., 2014) for constructing a linkage map with 130 F_2 samples total.

Due to the availability of samples, only 113 *C. hystrix* F_2 samples were sent for GBS (Elshire et al. 2011) at the Biotechnology Center, University of Wisconsin, Madison to increase the resolution of the genetic map. The samples were cut with *Ape*K I enzyme and the library was run on the single-ended Illumina Hi-Seq 2000 sequencing platform.

SNP calling and filtering

SNPs were called from raw GBS data consisting of 113 F_2 individuals using Tassel 5.0 program with default parameters (Glaubitz et al. 2014). *C. hystrix* TH1 draft genome was used as a reference. Plant individuals with fewer than 500,000 reads were removed from further analysis in the TASSEL 5.0 pipeline. After SNP calling, individuals with >50% missing genotypic data SNPs with more than 40% missing data points were filtered out. SNPs with minor allele

frequency (MAF) smaller than 0.05 were also excluded. The SNPs were further filtered for non-polymorphism between *C. hystrix* TH1 and CN1 parents.

Linkage map construction

The remaining high-quality SNPs were used to construct a linkage map using the function *mstmap* with Kosambi mapping function and p-value of 1e-9 in *ASMap* package in R software. The *mstmap* function is based on the minimum-spanning-tree (MST) algorithm and is fast and efficient for handling dense maps (Taylor & Butler, 2017; Wu, et al., 2008). A plot for pairwise recombination fraction (RF) by LOD score was created using the function *heatMap* in *R/ASMap* to examine the order of linked markers and segregation distortion.

Next, the SNP map was integrated with a previously developed genetic map with 410 SSR markers using the same *mstmap* function in *R/ASMap*. The integrated map was constructed with 74 F_2 individuals that have both complete SNP and SSR data. The integrated SNP-SSR map was used for further syntenic study among *C. hystrix*, cucumber and melon.

Syntenic relationships among the C. sativus, C. hystrix, and C. melo genomes

The *C. hystrix* TH1 scaffolds associated with SNP and SSR on the integrated linkage map were BLASTed against the cucumber Gy14 assembly version 2.0 (Xu et al., manuscript in preparation). The blast results were sorted based on their chromosome and Gy14 draft genome assembly positions to compare relative syntenic positions in the *C. hystrix* and cucumber genomes. *C. hystrix* TH1 scaffolds associated with SNPs on the linkage map were also blasted against the *C. melo* genome version 3.6.1 (Ruggieri et al. 2018) to infer syntenic relationship

among the three species.

Blast results and a karyotype for the *C. hystrix* genome previously developed by Dr. Dal-Hoe Koo (unpublished) were incorporated to refine the syntenic relationships among *C. hystrix*, cucumber and melon (*C. melo*, $2n = 2x = 24$) chromosomes. Syntenic blocks were defined based on *C. hystrix* chromosomes where breakages occurred when aligned to the cucumber or melon genome. The orientation of each syntenic block in relation to the Gy14 genome assembly was determined by the physical locations of Gy14 chromosomes when aligned to the *C. hystrix* genetic map.

Results

Linkage map construction of C. hystrix

1) SNP-based genetic map

A total of 194,600,926 out of 379,176,277 (51.3%) input raw reads were used to align against the *C. hystrix* TH1 assembly to call for SNPs. Two *C. hystrix* F_2 samples (#116 and #125) did not have any reads, leaving 111 F_2 samples for genetic map construction. In total, 16,675 SNPs were called from the TASSEL 5.0 pipeline. After filtering for missing data and non-polymorphic markers, 1,692 SNPs (98.8% genotyped and 1.2 % missing data) remained for linkage map construction. The map consisted of 12 chromosomes spanning 948.9 cM (**Table 3.1**) with an average spacing of 0.6 cM and anchored 1,069 scaffolds accounting for 49.4 Mb or about 21.85% of *C. hystrix* TH1 draft genome assembly (**Table 3.2**).

2) SSR-SNP-based consensus map

The SNP-based map was integrated with a previously developed genetic map with 410 SSR markers (Yang et al. 2014). The resulting consensus map consisted of 2,102 markers for 12 chromosomes spanning 1119 cM with an average spacing of 0.5 cM (**Table 3.1**). The SSR markers used in this map anchored 363 scaffolds, accounting for 19.0 Mb or 8.39% of the assembly. The integrated map anchored 1,321 scaffolds and accounted for 60.2 Mb or 26.62% of the assembly. Chromosome 2 had the largest number of markers which anchored the most scaffolds in total length of 8.89 Mb. Chromosome 10 had the least markers but anchored more base pairs from the assembly than chromosome 8. The statistics for anchored scaffolds was estimated from the clean version of *C. hystrix* TH1 draft genome assembly. The information regarding the genetic and physical locations of the syntenic blocks on *C. hystrix* and cucumber genomes, respectively, is presented in **Table 3.3**. SSR primer sequences can be found in Yang et al. (2014). The RF-LOD plots did not show obvious off-position of markers on both maps (**Figure 3.1**), indicating that the maps were of good quality for further syntenic study.

Segregation distortion analysis

SNP or SSR markers were tested for segregation distortion (SD) with χ^2 test and the p-value for each marker. *C. hystrix* is a diploid species with $n = 12$*;* thus, at least 24 independent genomic regions are expected – that is one possible recombination event per chromosome arm. A threshold of at least 0.05/24 ~0.002 (or -log10(p)~2.70) with Bonferroni correction was used to obtain a genome-wide error rate of alpha=0.05. From the consensus genetic map, no marker displayed segregation distortion with Bonferroni p-value > 2.70 (data not shown).

Syntenic relationships between C. hystrix and C. melo genomes

Blast results between *C. hystrix* scaffolds and *C. melo* ($2n = 2x = 24$) genome version 3.6.1 revealed that there was a reciprocal translocation between chromosomes 2 and 8. The rest of the genome was highly syntenic to melon with a few inversion events. This result was essentially consistent with Yang et al. (2014). As such, no further discussion was made on this topic.

Syntenic relationships between C. sativus and C. hystrix genomes

A karyotype for the *C. hystrix* genome was developed with fosmid markers by Dr. Dal-Hoe Koo (2013, unpublished). The karyotype and blast results between *C. hystrix* and cucumber genomes allowed for refinement of the syntenic relationships of chromosomes between the two specie. The results were largely consistent with findings made in a previous study (Yang et al. 2014) with 5 major and 8 minor structural changes. The first major difference suggested that there were 59 syntenic blocks as compared to 53 blocks from the previous study. Twenty-nine blocks (49.15%) were collinear and 30 blocks (50.15%) were inverted between *C. hystrix* and cucumber. The rest of changes involved combining syntenic blocks together in cucumber Chromosome C4, new syntenic blocks on *C. hystrix* chromosomes H03, H08, H09 and H12, one missing block in *C. hystrix* chromosomes H06 and H10, and the change in centromere location of cucumber chromosome C6 (see below for discussion).

The inferred syntenic relationship between these two species could be summarized as follows: C1 = H02 + H08 + H12, C2 = H03 + H05 + H11, C3 = H04 + H06, C4 = H05 + H07 + H08, C5 = H09 + H10, C6 = H02 + H03 + H05 + H08 + H11, and C7 = H01 where C represented cucumber chromosomes and H represented *C. hystrix* chromosomes (**Table 3.4**). The entire

chromosomal rearrangement involved 4 translocations, 4 chromosomal fusions, 18 pericentric inversions and 35 paracentric inversions. An improved model to elucidate the mechanisms of dysploid reduction from n = 12 in the common ancestor to n = 7 in cucumber is proposed below.

Cucumber chromosome 7

Blast results showed that C7 was completely syntenic to H01 with an inversion in the short arm of C7 because the marker order on the *C. hystrix* genetic map was inverted between 0 – 51.23 cM when compared to the cucumber Gy14 genome assembly (**Table 3.3**). The inversion was 5 Mb, in size spanning from 0-5Mb of C7. The result was consistent with the karyotype of H01, where the locations of fosmid loci 71.1-71.6 were inverted (**Figure 3.2B, H1**) as compared to C7 (Figure 5 in Yang et al., 2014). Fosmid loci 71.7 and 7.18 on cucumber chromosome C7 karyotype were also located in the opposite order of those in H1. Fosmid loci 71.9-72.4 and blast results of the map were completely collinear for the rest of C7. The karyotype results indicated that there were probably two inversion events. Thus, the proposed model for the evolution of ancient chromosome 1 (AK1) of the common ancestor to C7 was the same as proposed by Yang et al. (2014). AK1 had broken into three syntenic blocks 1B1-1B3 and undergone two inversion events, pericentric followed paracentric inversion, to become C7 (see Figure S5A in Yang et al., 2014).

Cucumber chromosome 3

Blast results showed that C3 genomic DNA sequences were highly similar to H4 and H6 with one chromosome fusion, 3 pericentric inversion and 4 paracentric inversion events. The proposed model was that AK4 had 5 syntenic blocks and AK6 had 3 syntenic blocks instead of 4 blocks as

in the previous study (Figure S5C in Yang et al., 2014). Both AK4 and AK6 had independently undergone pericentric inversions before chromosome fusion. Then, the inverted chromosomes fused and underwent one more round of pericentric inversion and four rounds of paracentric inversion to become cucumber chromosome 3 (**Figure 3.3A**). Syntenic block 4B3 was located at 59.42 cM region on the *C. hystrix* genetic map with no recombination and was in the pericentromeric region (between fosmid loci 32.2-32.5) according to the karyotype of chromosome 4 (**Figure 3.2B, H4**), thus the orientation of the block when aligned to cucumber could not be determined.

Cucumber chromosome 5

Blast results showed that cucumber chromosome C5 was completely syntenic to *C. hystrix* chromosomes H09 and H10. AK9 and AK10 fused with each other and followed by a series of inversion events to become cucumber chromosome 5 (**Figure 3.3B**). There were 8 syntenic blocks of AK9 instead of 4 blocks and 2 syntenic blocks of AK10 instead of 3 blocks. The model included 10 chromosomal rearrangement events: 1 fusion, 3 pericentric inversion and 6 paracentric inversion events instead of 1 fusion, 2 pericentric inversion and 8 paracentric inversion in the previous study (Figure S5D in Yang et al., 2014). The proposed model resulted in fewer chromosomal rearrangements despite having more syntenic blocks. This was probably due to the much improved cucumber genome assembly (Gy14 assembly version 2.0) and the higher density linkage map we used in this study. .

Cucumber chromosomes 1, 4, 2 and 6

Shubert and Lysak (2011) suggested that a more parsimonious path of chromosome

rearrangements is more likely. Blast results and the karyotypes of *C. hystrix* suggested that cucumber chromosomes 7, 3, and 5 arise independently. However, cucumber chromosomes 1, 4, 2, and 6 must arise in this order since they accounted for the least number of chromosomal rearrangement events, including translocations, fusions and inversions, from *n*=12 chromosomes in the common ancestor to *n*=7 chromosomes in cucumber (**Figure 3.4**).

Firstly, AK2 and AK8 underwent symmetric reciprocal translocation resulting in AK2/8(i) and AK2/8(ii) intermediate chromosomes. Then, the AK2/8(ii) fused with AK12 in a process called symmetric nested fusion by reciprocal translocation (Schubert and Lysak 2011) followed by 3 pericentric and 3 paracentric inversion events to become cucumber chromosome 1 (**Figure 3.3C**).

The AK2/8(i) underwent symmetric reciprocal translocation with AK5 to form AK2/5/8 and AK 5/8 intermediate chromosomes. AK5/8 fused with AK7 by a series of events including 1 more translocation, 1 fusion, 3 pericentric inversion and 7 paracentric inversion events to become cucumber chromosome 4 (**Figure 3.3D**).

The AK2/5/8 underwent symmetric reciprocal translocation with AK11 and formed AK5/11 and AK2/5/8/11 intermediate chromosomes. AK5/11 fused with AK3 by asymmetrical translocation and formed a short part of AK3 and AK3/5/11. AK3/5/11 underwent 3 more pericentric inversion and 6 paracentric inversion events to form cucumber chromosome 2 (**Figure 3.3E**). Finally, the part of AK3 that remained after asymmetrical translocation fused with AK2/5/8/11 followed by 2 pericentric inversion and 8 paracentric inversion events to form cucumber

chromosome 6 (**Figure 3.3F**).

Discussion

Comparison of the current genetic map with previous C. hystrix map.

The current integrated SNP-SSR genetic map anchored 1,321 scaffolds accounting for 60.2 Mb or about 26.62% of *C. hystrix* TH1 draft genome assembly. The previous genetic map was constructed from an F_2 population with 91 individuals and was consisted of only 417 SSR markers, which anchored 197 scaffolds (328 kb total) and 109 contigs (110 kb total) or about 0.21% of the previous *C. hystrix* genome assembly. The improvement from the previous study was due to the low cost of genotyping sequencing technology and availability of SNP markers that led to the new genetic map with higher resolution. The current map also provided a more detailed insight into the syntenic relationship between *C. hystrix* and cucumber genomes.

Marker order on C. hystrix genetic map

C. hystrix chromosome 2

The fosmid loci 11.8 and 11.9 on the *C. hystrix* genetic map (**Table 3.3**) were in the opposite order as compared to their locations in the karyotype of *C. hystrix* chromosome 2 (239G12 and 60D15 in **Figure 3.2A, H2**, respectively). This was probably because both fosmid loci were in the heterochromatic region near the centromere with low or no recombination. Thus, markers placed in this region could be in an incorrect order on the genetic map.

C. hystrix chromosome 3

When *C. hystrix* scaffolds anchored on the genetic map were aligned against cucumber Gy14

genome, no markers were found in one region spanning 7 Mb in cucumber chromosome 2 from 20.9-27.0 Mb (**Table 3.3**). This region was highly heterochromatic in cucumber chromosome 2 (see Figure 3 in Yang et al., 2012). This region corresponded to the region in *C. hystrix* chromosome 3 between fosmid loci 62.5 and 22.0 (44K19 and 70O3 in **Figure 3.2A, H3** respectively), which also appeared to be highly heterochromatic.

C. hystrix chromosome 11

Markers on the *C. hystrix* genetic map within the syntenic block 11B3 (26.42 – 45.63 cM) might be disordered and the orientation of the syntenic block 11B3 could not be correctly determined. Alignment result in this region was not collinear with positions on Gy14 chromosome 6 . Syntenic block 11B3 was in a pericentromeric region of chromosome 11 between fosmid loci 62.0 and 62.2 (8C13 and 59D13 in **Figure 3.2A, H11**, respectively). Block 11B3 was also syntenic to the centromeric region in cucumber chromosome 6.

Conservation of cucumber chromosome 7 during the evolution of Cucumis

C7 was highly conserved and showed the highest level of synteny to H1 and melon chromosome 1 as compared to other cucumber chromosomes (Yang et al., 2014). C7 differed from H1 and melon chromosome 1 only by two inversion events in the short arm of the chromosome. However, the region in the syntenic block 1B2, spanning from 7.5-16 Mb in C7, appeared to have no recombination at 51.91 cM in the *C. hystrix* genetic map. The karyotype results indicated that there were probably two inversion events despite no recombination in the block 1B2. The lack of recombination in this region was probably due to the small population size used to construct the map and that a larger population size should be used in the future.

Genomic rearrangements in dysploid reduction from n=12 to n=7

The proposed model in this study, in which ancestor chromosomes evolved from *n*=12 to *n*=7 in cucumber, included 39 paracentric inversions and 18 pericentric inversions. The prevalence of paracentric inversions as compared to pericentric inversions was supported by the hypothesis that these inversions have fewer harmful effects on fertility, and therefore fitness, of a new species than pericentric inversions (Frary *et al.*, 2016). Recurrent chromosomal breakpoints, commonly found on pericentromeric regions, were primarily responsible for genome restructuring in tomato (Wu *et al.*, 2010). While it was unclear of recurrent breakpoint locations on chromosomes of the common ancestor between *C. hystrix* and cucumber, we observed four breakpoints followed by pericentric inversions on AK1, AK3, AK4, and AK6 in the proposed model.

Acknowledgement

I would like to thank Dr. Luming Yang for *C. hystrix* F_2 population development and partial work on F_2 genotyping with SSR markers (Yang et al., 2014). I thank Dr. Dal-Hoe Koo for developing karyotype for *C. hystrix* with fosmid markers (unpublished). I thank Kristin Haider for technical help, Dr. Douglas Senalik for providing a GBS SNP-calling pipeline.

References

Chen J, Moriarty G, Jahn M (2004) Some disease resistance tests in *Cucumis hystrix* and its progenies from interspecific hybridization with cucumber. In: Progress in cucurbit genetics and breeding research. Proceedings of Cucurbitaceae 2004, the 8th EUCARPIA meeting on Cucurbit Gentics and Breeding. pp 189–196

Chen J, Staub JE, Tashiro Y, et al (1997) Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr. Euphytica 96:413–419. doi: 10.1023/A:1003017702385

Chen J, Zhou XH (2011) Wild Crop Relatives: Genomic and Breeding Resources (Vegetables). In: Chittaranjan K (ed) Wild Crop Relatives: Genomic and Breeding Resources Vegetables. Springer, pp 67–87

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS One 6:e19379. doi: 10.1371/journal.pone.0019379

Frary A, Doganlar S, Frary A (2016) Synteny Among Solanaceae Genomes. In: The Tomato Genome. Compendium of Plant Genomes. Springer, Berlin, Heidelberg. Springer, Berlin, Heidelberg, pp 217–243

Glaubitz JC, Casstevens TM, Lu F, et al (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346. doi: 10.1371/journal.pone.0090346

Han YH, Zhang ZH, Liu JH, et al (2008) Distribution of the tandem repeat sequences and karyotyping in cucumber (*Cucumis sativus* L.) by fluorescence in situ hybridization. Cytogenet Genome Res 122:80–8. doi: 10.1159/000151320

Jiang B, Wu Z, Lou Q, et al (2010) Genetic diversity of Ty1-*copia* retrotransposons in a wild species of Cucumis (*C. hystrix*). Sci Hortic (Amsterdam) 127:46–53. doi: 10.1016/j.scienta.2010.09.013

Li K, Wang H, Wang J, et al (2016) Divergence between *C. melo* and African *Cucumis* Species Identified by Chromosome Painting and rDNA Distribution Pattern. Cytogenet Genome Res. doi: 10.1159/000453520

Li Y, Yang L, Pathak M, et al (2011) Fine genetic mapping of *cp*: a recessive gene for compact (dwarf) plant architecture in cucumber, *Cucumis sativus* L. Theor Appl Genet 123:973–83. doi: 10.1007/s00122-011-1640-6

Murray MG, Thompson WF (1980) Nucleic Acids Research Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4326. doi: doi:10.1093/nar/8.19.4321

Ruggieri V, Alexiou KG, Morata J, et al (2018) An improved assembly and annotation of the melon (*Cucumis melo* L.) reference genome. Sci Rep 8:8088. doi: 10.1038/s41598-018-26416-2

Schubert I, Lysak MA (2011) Interpretation of karyotype evolution should consider chromosome structural constraints. Trends Genet 27:207–216. doi: 10.1016/j.tig.2011.03.004

Taylor J, Butler D (2017) R Package ASMap: Efficient Genetic Linkage Map Construction and Diagnosis. J Stat Softw 79:1–29. doi: 10.18637/jss.v079.i06

Wu F, Eannetta NT, Xu Y, et al (2010) COSII genetic maps of two diploid *Nicotiana* species provide a detailed picture of synteny with tomato and insights into chromosome evolution in tetraploid *N. tabacum*. Theor Appl Genet 120:809–827. doi: 10.1007/s00122-009-1206-z

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and Accurate Construction of Genetic Linkage Maps from the Minimum Spanning Tree of a Graph. PLOS Genet 4:e1000212. doi: 10.1371/JOURNAL.PGEN.1000212

Yang L, Koo D-H, Li Y, et al (2012) Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. Plant J 71:895–906. doi: 10.1111/j.1365-313X.2012.05017.x

Yang L, Koo DH, Li D, et al (2014) Next-generation sequencing, FISH mapping and synteny-based modeling reveal mechanisms of decreasing dysploidy in *Cucumis*. Plant J 77:16–30. doi: 10.1111/tpj.12355

Yang S, Qin X, Cheng C, et al (2017) Organization and evolution of four differentially amplified tandem repeats in the *Cucumis hystrix* genome. Planta 246:749–761. doi: 10.1007/s00425-017-2716-6

Zhang Y, Cheng C, Li J, et al (2015) Chromosomal structures and repetitive sequences divergence in *Cucumis* species revealed by comparative cytogenetic mapping. BMC Genomics 16:730. doi: 10.1186/s12864-015-1877-6

Table 3.1 Summary of *C. hystrix* SNP linkage map and an integrated SNP-SSR map. The number of anchored scaffolds and total length in bp were calculated from the integrated map and *C. hystrix* TH1 assembly (clean version).

Chr	SNP map				SNP + SSR map					
	No. markers	Map Length (cM)	Average map interval (cM)	Maximum map interval (cM)	No. markers	Map Length (cM)	Average map interval (cM)	Maximum map interval (cM)	No. anchored scaffolds	Total length (Mb)
H01	126	93.8	0.8	8.1	163	109.7	0.7	7.6	112	5.52
H02	247	108.3	0.4	6.7	293	127.1	0.4	5.2	178	8.89
H03	149	80.0	0.5	6.2	184	104.5	0.6	6.5	115	5.32
H04	206	100.1	0.5	6.4	238	126.0	0.5	6.1	148	5.88
H05	148	77.1	0.5	7.9	193	97.9	0.5	5.8	106	5.51
H06	110	96.2	0.9	10.8	145	111.4	0.8	9.9	95	4.75
H07	157	77.0	0.5	7.3	189	92.6	0.5	8.5	117	5.15
H08	98	52.9	0.5	5	122	67.1	0.6	4.3	83	3.05
H09	125	68.5	0.6	6.2	151	81.2	0.5	6.1	93	4.53
H10	82	52.4	0.6	8.8	111	56.2	0.5	7.1	68	3.47
H11	118	75.8	0.6	4.9	157	79.2	0.5	4.5	109	4.58
H12	126	66.8	0.5	7.9	156	66.1	0.4	8.9	97	3.52
Total	**1,692**	**948.9**	**0.6**	**10.8**	**2,102**	**1,119.0**	**0.5**	**9.9**	**1,321**	**60.16**

Table 3.2 Summary of *C. hystrix* TH1 genome assembly that was anchored by the genetic map.

Genetic map	**No. markers**	**No. scaffolds**	**Total length (bp)**	**% assembly***
SNP map	1,662	1,069	49,374,712	21.85
SSR map	413	363	18,968,736	8.39
SNP-SSR map	2,102	1,321	60,158,887	26.62

* Estimated from the clean version of *C. hystrix* TH1 draft genome assembly

Table 3.3 Summary of syntenic relationship between *C. hystrix* and cucumber genomes. Physical positions of cucumber are based on cucumber Gy14 assembly version 2.0. Number in front "B" indicates ancestral chromosome and number after "B" indicates syntenic block number. +/- refers to the orientation of syntenic blocks relative to cucumber Gy14 genome physical position.

#	**Gy14 chr**	**Gy14 pos (Mb)**	**Syntenic block no.**	**Syntenic block orientation**	***C. hystrix* chr**	***C. hystrix* pos (cM)**
1	1	0 - 2.7	2B4	-	H02	102.89 - 127.07
2	1	2.9 - 3.0	12B4	+	H12	66.12
3	1	3.2 - 11.8	12B1	+	H12	0 - 45.69
4	1	13.8 - 19.2	2B3	+	H02	68. 40 - 98.00
5	1	19.7 - 23.4	12B3	-	H12	47.7 - 66.12
6	1	23.9 - 24.8	8B6	-	H08	58.9 - 67.08
7	1	24.8 - 31.6	2B2	+	H02	54.88 - 68.40
8	1	32.5 - 33.2	12B2	-	H12	47.04 - 47.72
9	2	0 - 9.2	5B6	-	H05	30.32 - 97.89
10	2	9.6 - 11.1	5B4	+	H05	15.58 - 25.46
11	2	11.8 - 16.1	11B1	-	H11	0 - 21.95
12	2	20.9 - 27.2	3B5	+	H03	53.9 - 55.3
13	2	27.2 - 31.7	3B6	-	H03	58.2 - 104.47
14	2	32.1 - 32.8	3B1	+	H03	0 - 9.98
15	2	34.1 - 34.7	11B2	-	H11	24.8 - 26.4
16	3	0 - 6.0	6B1	+	H06	0 - 39.97
17	3	6.4 - 19.2	6B3	-	H06	40.78 - 111.39
18	3	19.4 - 21.6	4B3	-	H04	57.93 - 59.42
19	3	23.5 - 25.3	6B2	+	H06	40.78
20	3	25.5 - 27.9	4B2	-	H04	55.77 - 59.42
21	3	28.0 -28.9	4B4	+	H04	60.91 - 63.07
22	3	29.4 - 36.2	4B5	-	H04	65.91 - 125.30
23	3	36.2 - 41.4	4B1	-	H04	2.16 - 55.77
24	4	0 - 8.8	7B4	-	H07	34.25 - 91.13
25	4	8.9 - 10.2	8B3	+	H08	2.16 - 6.49
26	4	10.4 - 10.6	7B5	+	H07	91.89 - 92.56
27	4	11.7 - 11.8	5B1	-	H05	0
28	4	17.8 - 18.5	5B2	+	H05	0.68 - 3.65
29	4	18.6 - 20.9	8B4	-	H08	7.16 - 41.93
30	4	20.9 - 21.5	8B1	-	H08	1.49
31	4	22.2 - 22.5	7B3	-	H07	34.92
32	4	22.9 - 26.8	8B2	+	H08	1.49

#	Gy14 chr	Gy14 pos (Mb)	Syntenic block no.	Syntenic block orientation	*C. hystrix* chr	*C. hystrix* pos (cM)
33	4	27.2 - 27.8	7B2	-	H07	30.96
34	4	27.8 - 31.1	7B1	+	H07	0 - 30.96
35	5	0 - 3.6	9B8	-	H09	41.10 - 81.16
36	5	4.0 - 13.2	10B2	+	H10	44.14 - 56.21
37	5	14.9 - 15.0	9B4	+	H09	27.04 - 27.71
38	5	16.8 - 18.5	9B6	-	H09	28.39 - 29.74
39	5	18.9 - 20.8	9B3	+	H09	26.36 - 28.39
40	5	21.4 - 21.8	9B5	+	H09	27.71 - 28.39
41	5	22.0 -24.1	9B2	-	H09	18.84 - 26.36
42	5	25.0 - 26.0	9B7	+	H09	31.90 - 39.07
43	5	26.1 - 30.8	10B1	-	H10	0 - 43.33
44	5	31.4 - 33.3	9B1	-	H09	0 - 16.0
45	6	0 - 0.7	11B6	+	H11	53.87 - 58.61
46	6	0.8 - 4.2	3B2	+	H03	9.99 - 36.27
47	6	4.7 - 4.8	11B4	+	H11	46.30 - 46.98
48	6	4.9 - 7.4	11B7	-	H11	60.77 - 79.17
49	6	7.6 - 15.2	11B3	+	H11	28.45 - 43.60
50	6	15.5 - 18.0	11B5	+	H11	47.65 - 51.85
51	6	18.2 - 20.6	3B3	+	H03	38.97 - 51.95
52	6	21.0 -21.5	3B4	-	H03	52.63 - 53.98
53	6	21.7 - 22.6	5B5	+	H05	25.45 - 28.97
54	6	22.7 - 23.7	5B3	-	H05	4.32 - 15.57
55	6	23.7 - 25.0	8B5	+	H08	42.61 - 58.97
56	6	25.1 - 32.2	2B1	-	H02	0 - 53.40
57	7	0 - 5.1	1B1	-	H01	0 - 51.23
58	7	7.5 - 16.1	1B2	-	H01	51.91
59	7	16.8 - 23.7	1B3	+	H01	53.40 - 109.66

Table 3.4 Summary of chromosomal rearrangement events during dysploid reduction from the *n*=12 ancestor to *n*=7 in cucumber.

Cucumber chromosome	Ancestor donor chromosome[a]	No. Syntenic blocks	No. Translocations	No. Fusions	No. Inversions[c]	Centromere inherited from
1	AK2/8 (2)[b], AK12	8	1	1	6 (3)	AK12
2	AK5/11, AK3	7	2	-	9 (3)	AK5
3	AK4, AK6	8	-	1	7 (3)	AK4 or AK6
4	AK5/8, AK7	11	2	1	10 (3)	AK7 or AK8
5	AK9, AK10	10	-	1	9 (3)	AK9
6	AK2/5/8/11, AK3	12	-	1	10 (2)	AK11
7	AK1	3	-	-	2 (1)	AK1
Total	-	59	4[d]	4	53	-

a. The ancestral karyotype is similar to or the same as the *C. hystrix* karyotype. AK1-AK12 correspond to H01-H12, respectively, in terms of chromosome structure and organization.
b. AK2/8 = result of the chromosome translocation between AK2 and AK8 that preceded all other events.
c. The numbers in parentheses indicate pericentric inversion.
d. Four translocations in total because C2 and C4 were counted independently of each other.

Figure 3.1 Pairwise recombination fractions (upper left diagonal) and LOD scores (lower right diagonal) for all markers. A) SNP marker order generated by *R/ASMap* based on 12 linkage groups of *C. hystrix* genome; B) Integrated consensus map with SSR and SNP markers generated by *R/ASMap.* X and Y axes are markers on each chromosome. Scale bar on left represents recombination fraction.

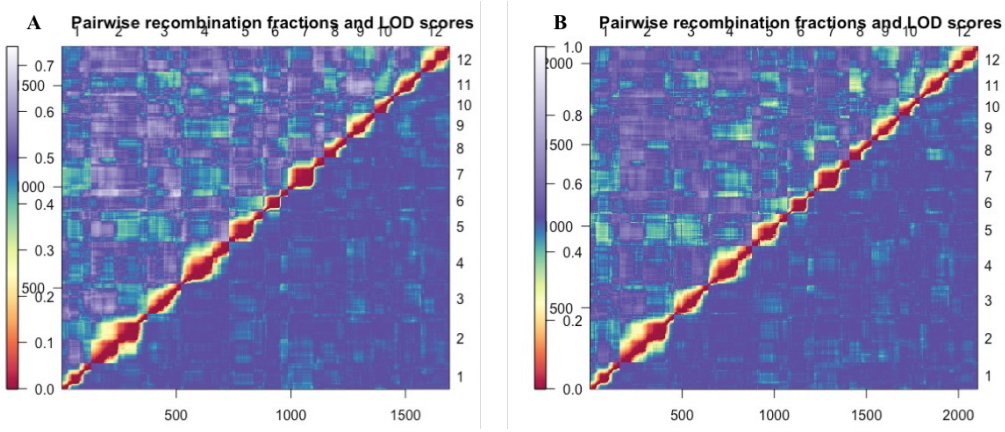


Figure 3.2 Karyotype for 12 *C. hystrix* chromosomes. (A) Pachytene FISH analysis was performed with 128 fosmid probes (in red and green colors). Arrows indicate estimated centromere locations. (B) A simplified cartoon version detailing fosmid location (white stripe) on each chromosome. Black rectangular shape corresponds to centromere, black circle at chromosome end corresponds to heterochromatic knob, NOR: nucleolar organizing region; dark grey color around the centromere corresponds to heterochromatic region (Koo, D., unpublished).

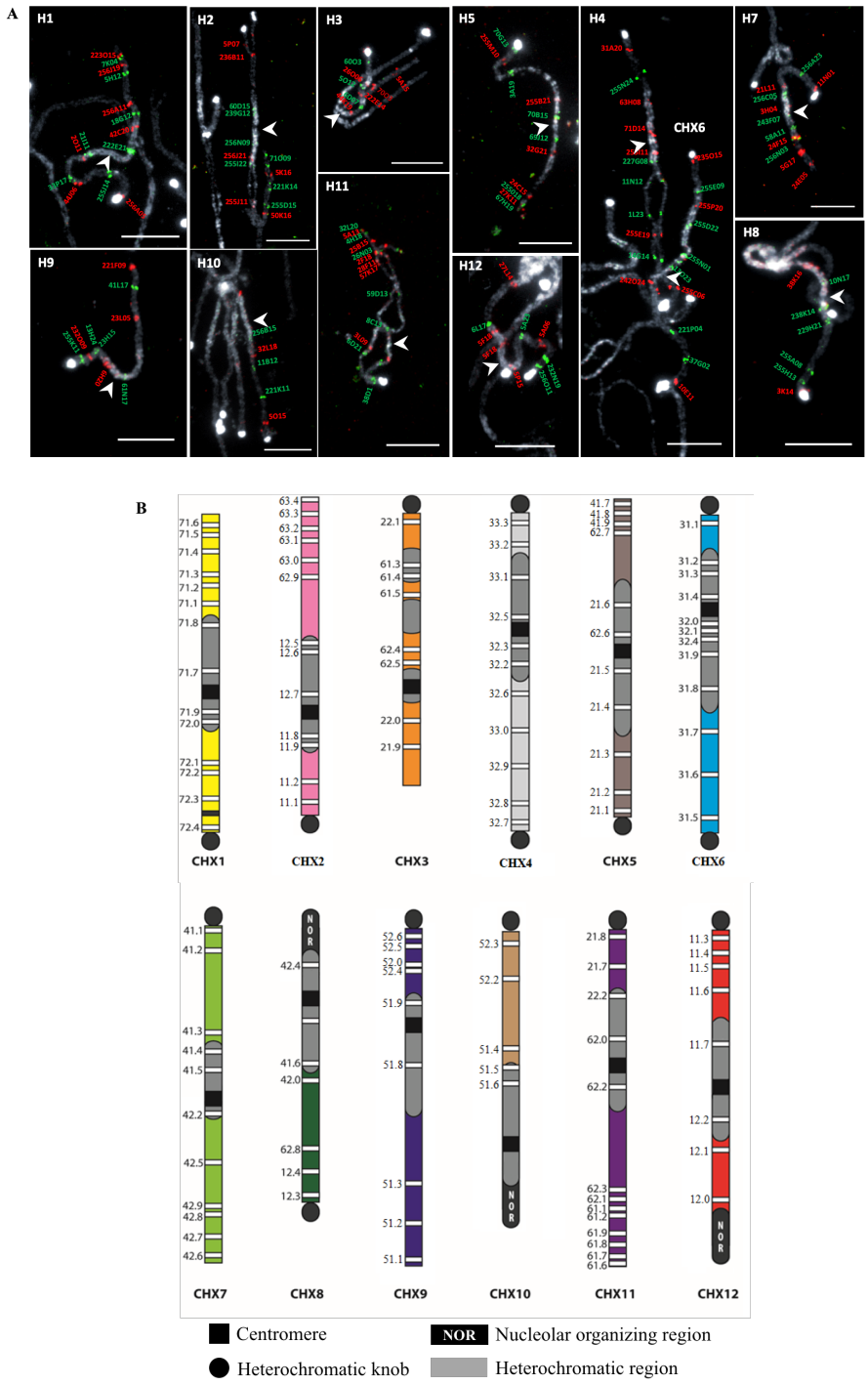


Figure 3.3 Reconstruction of karyotype evolution history of seven cultivated cucumber chromosomes 1 to 7 from 12 ancestral chromosomes AK1-AK12. Number in front "B" indicates ancestral chromosome and number after "B" indicates syntenic block number. +/- refers to the orientation of syntenic blocks relative to cucumber Gy14 genome physical position. A) cucumber chromosome 3, B) cucumber chromosome 5, C) cucumber chromosome 1, D) cucumber chromosome 4, E) cucumber chromosome 2, and F) cucumber chromosome 6. Ipe refers to pericentric inversion, Ipa paracentric inversion, T translocation, and F chromosome fusion.

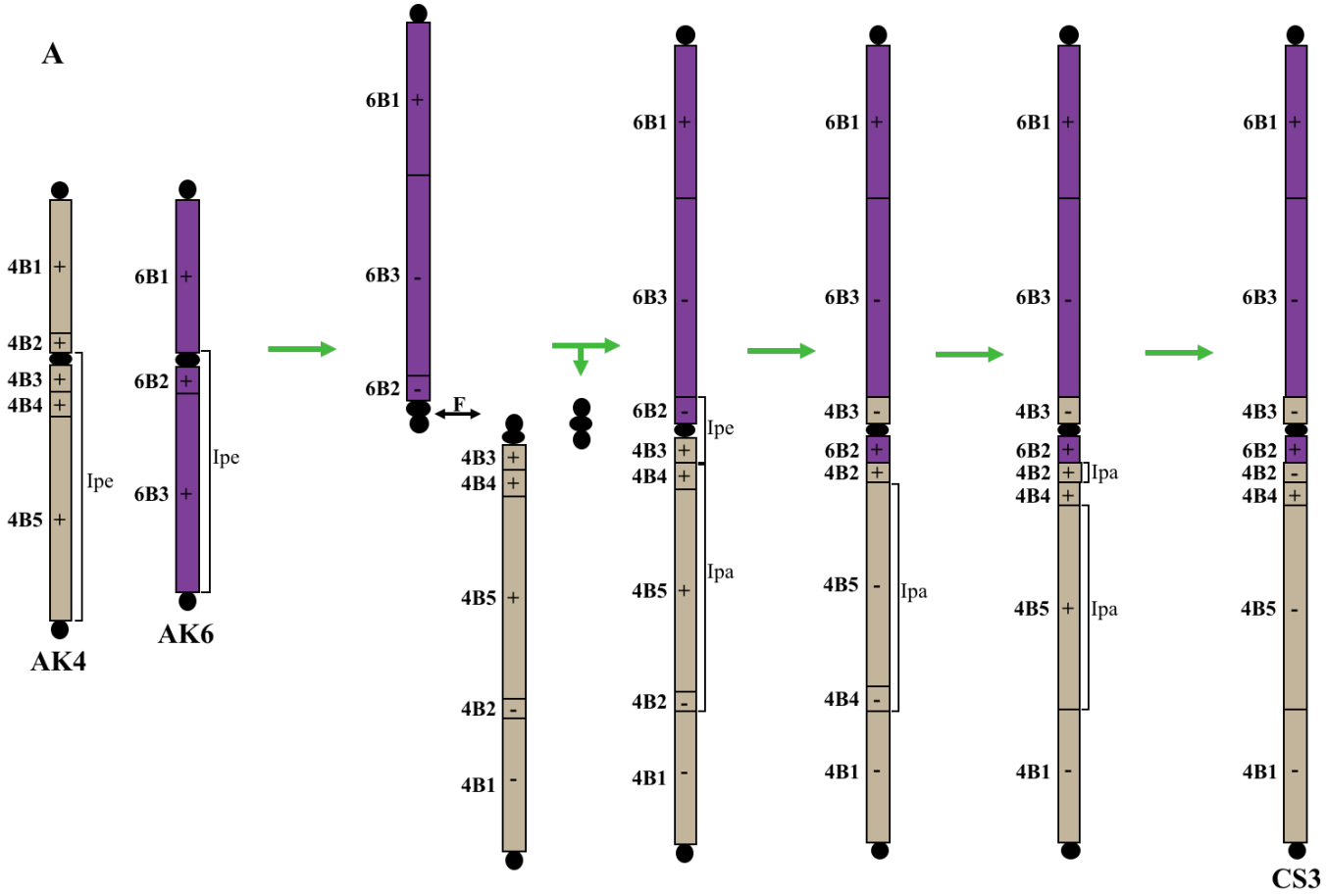


Figure 3.3 Reconstruction of karyotype evolution history of seven cultivated cucumber chromosomes 1 to 7 from 12 ancestral chromosomes AK1-AK12. Number in front "B" indicates ancestral chromosome and number after "B" indicates syntenic block number. +/- refers to the orientation of syntenic blocks relative to cucumber Gy14 genome physical position. A) cucumber chromosome 3, B) cucumber chromosome 5, C) cucumber chromosome 1, D) cucumber chromosome 4, E) cucumber chromosome 2, and F) cucumber chromosome 6. Ipe refers to pericentric inversion, Ipa paracentric inversion, T translocation, and F chromosome fusion.

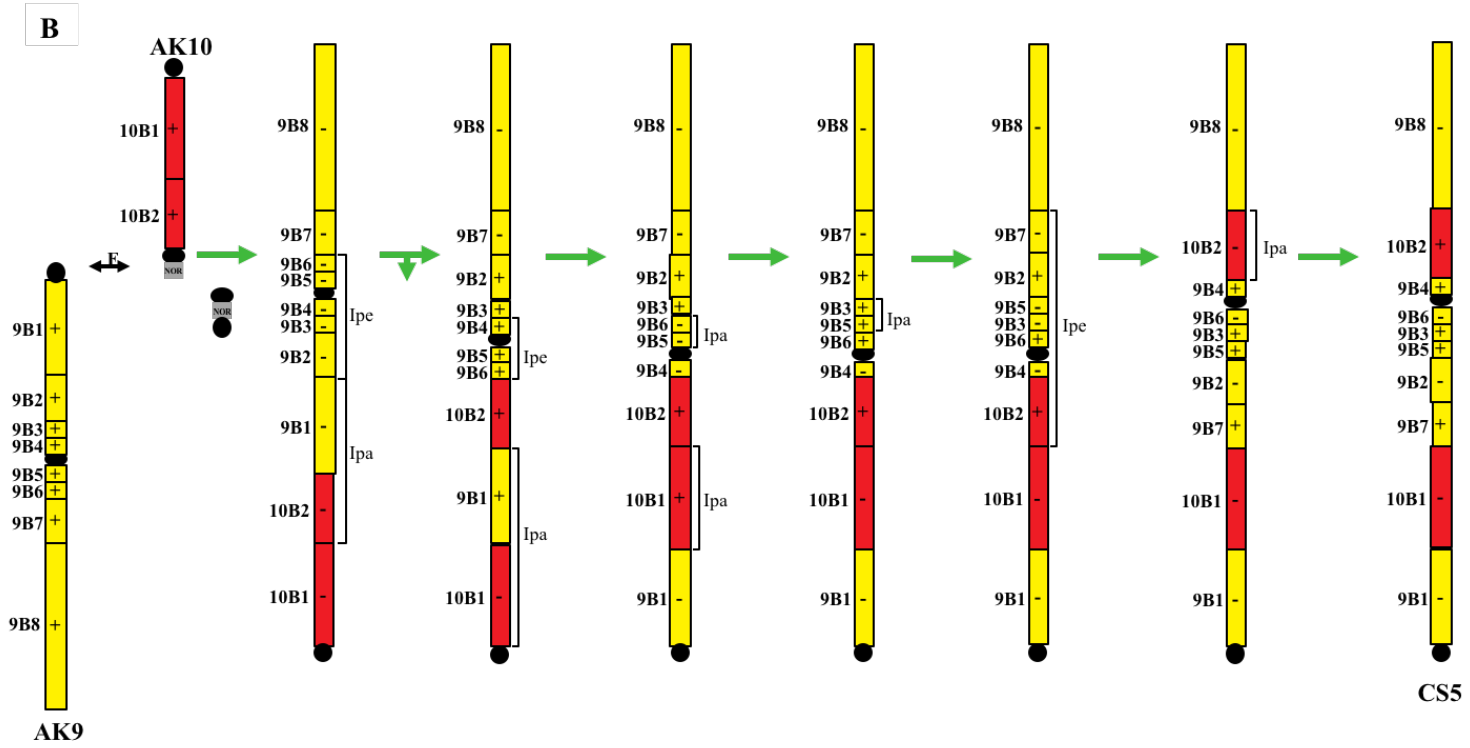


Figure 3.3 Reconstruction of karyotype evolution history of seven cultivated cucumber chromosomes 1 to 7 from 12 ancestral chromosomes AK1-AK12. Number in front "B" indicates ancestral chromosome and number after "B" indicates syntenic block number. +/- refers to the orientation of syntenic blocks relative to cucumber Gy14 genome physical position. A) cucumber chromosome 3, B) cucumber chromosome 5, C) cucumber chromosome 1, D) cucumber chromosome 4, E) cucumber chromosome 2, and F) cucumber chromosome 6. Ipe refers to pericentric inversion, Ipa paracentric inversion, T translocation, and F chromosome fusion.

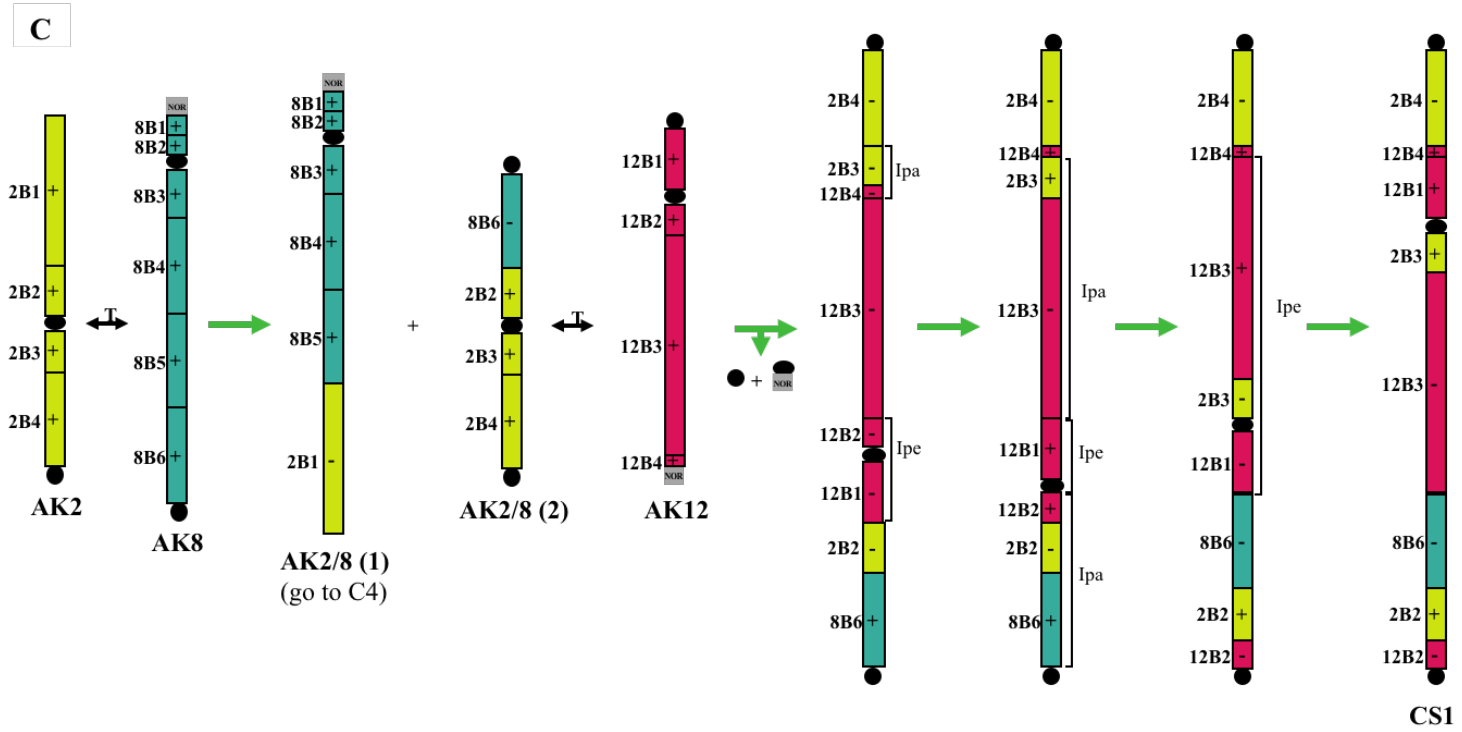


Figure 3.3 Reconstruction of karyotype evolution history of seven cultivated cucumber chromosomes 1 to 7 from 12 ancestral chromosomes AK1-AK12. Number in front "B" indicates ancestral chromosome and number after "B" indicates syntenic block number. +/- refers to the orientation of syntenic blocks relative to cucumber Gy14 genome physical position. A) cucumber chromosome 3, B) cucumber chromosome 5, C) cucumber chromosome 1, D) cucumber chromosome 4, E) cucumber chromosome 2, and F) cucumber chromosome 6. Ipe refers to pericentric inversion, Ipa paracentric inversion, T translocation, and F chromosome fusion.

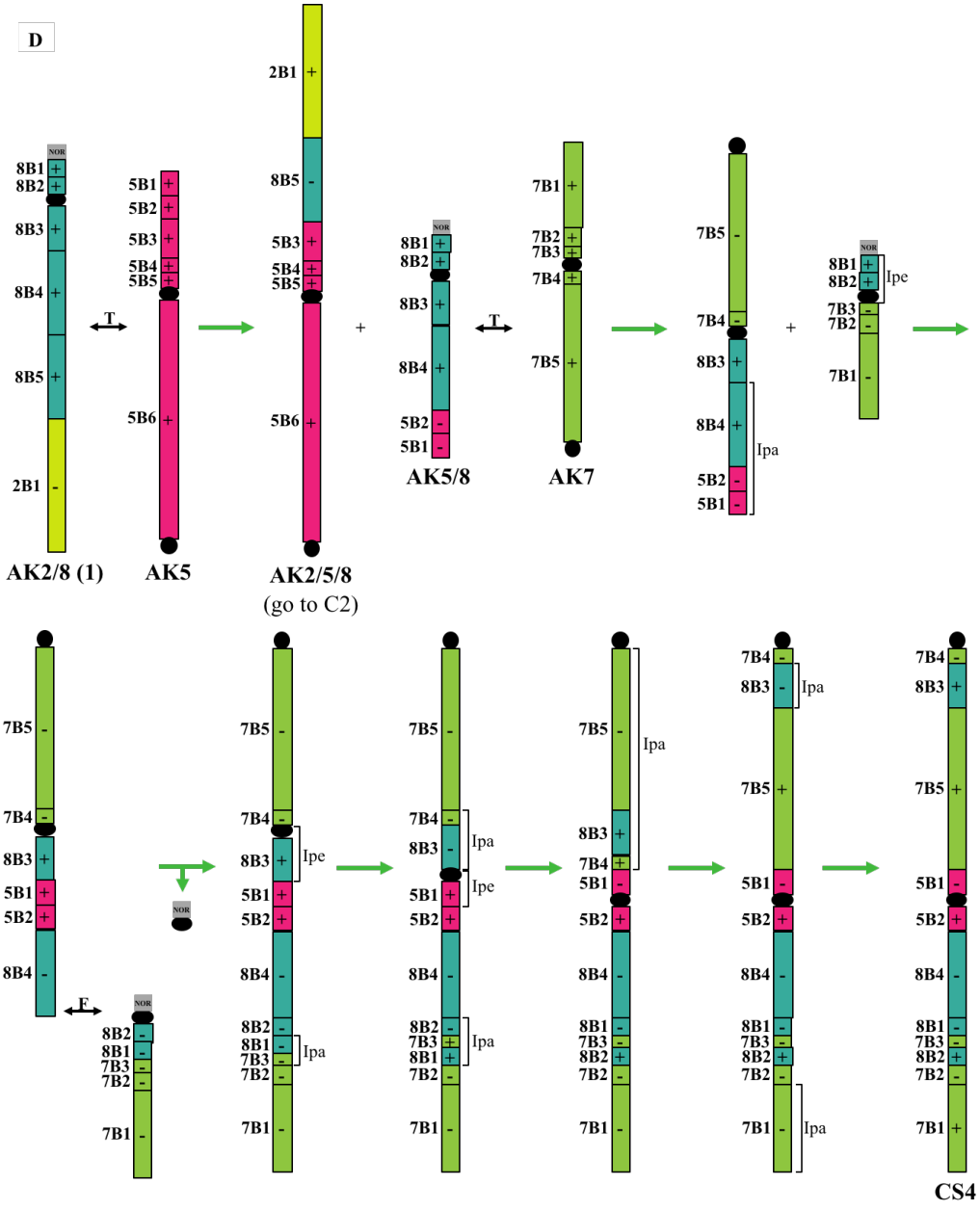


Figure 3.3 Reconstruction of karyotype evolution history of seven cultivated cucumber chromosomes 1 to 7 from 12 ancestral chromosomes AK1-AK12. Number in front "B" indicates ancestral chromosome and number after "B" indicates syntenic block number. +/- refers to the orientation of syntenic blocks relative to cucumber Gy14 genome physical position. A) cucumber chromosome 3, B) cucumber chromosome 5, C) cucumber chromosome 1, D) cucumber chromosome 4, E) cucumber chromosome 2, and F) cucumber chromosome 6. Ipe refers to pericentric inversion, Ipa paracentric inversion, T translocation, and F chromosome fusion.

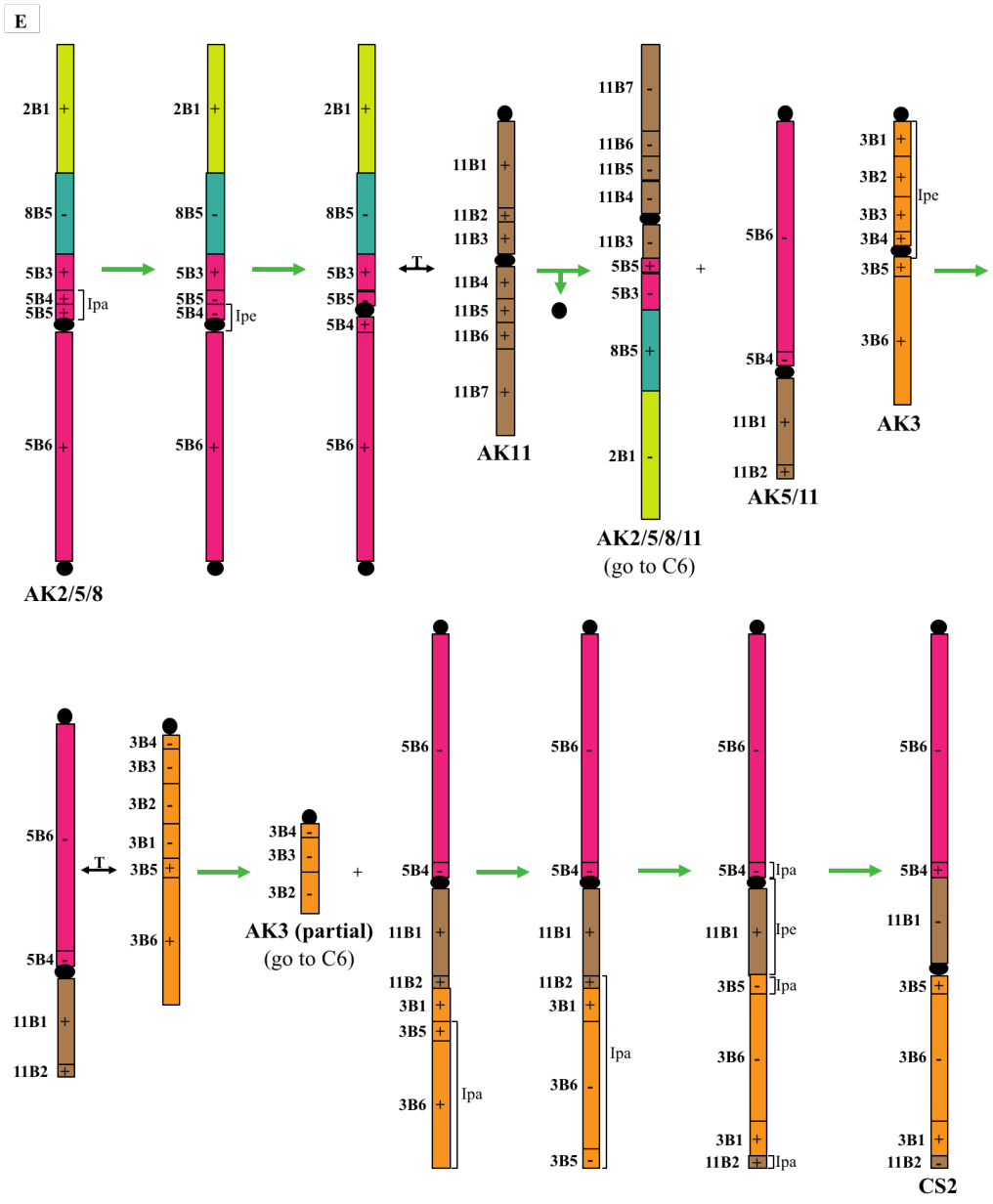


Figure 3.3 Reconstruction of karyotype evolution history of seven cultivated cucumber chromosomes 1 to 7 from 12 ancestral chromosomes AK1-AK12. Number in front "B" indicates ancestral chromosome and number after "B" indicates syntenic block number. +/- refers to the orientation of syntenic blocks relative to cucumber Gy14 genome physical position. A) cucumber chromosome 3, B) cucumber chromosome 5, C) cucumber chromosome 1, D) cucumber chromosome 4, E) cucumber chromosome 2, and F) cucumber chromosome 6. Ipe refers to pericentric inversion, Ipa paracentric inversion, T translocation, and F chromosome fusion.

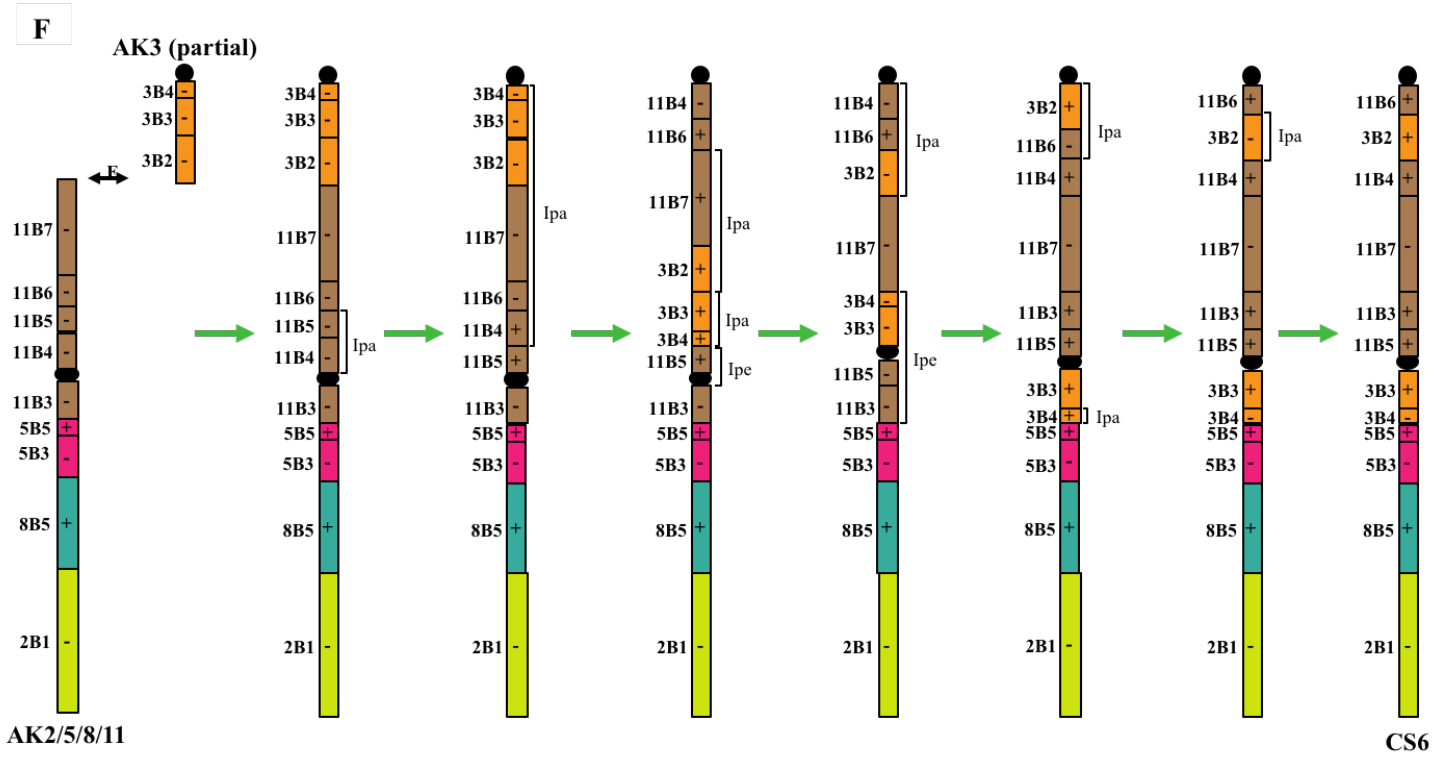
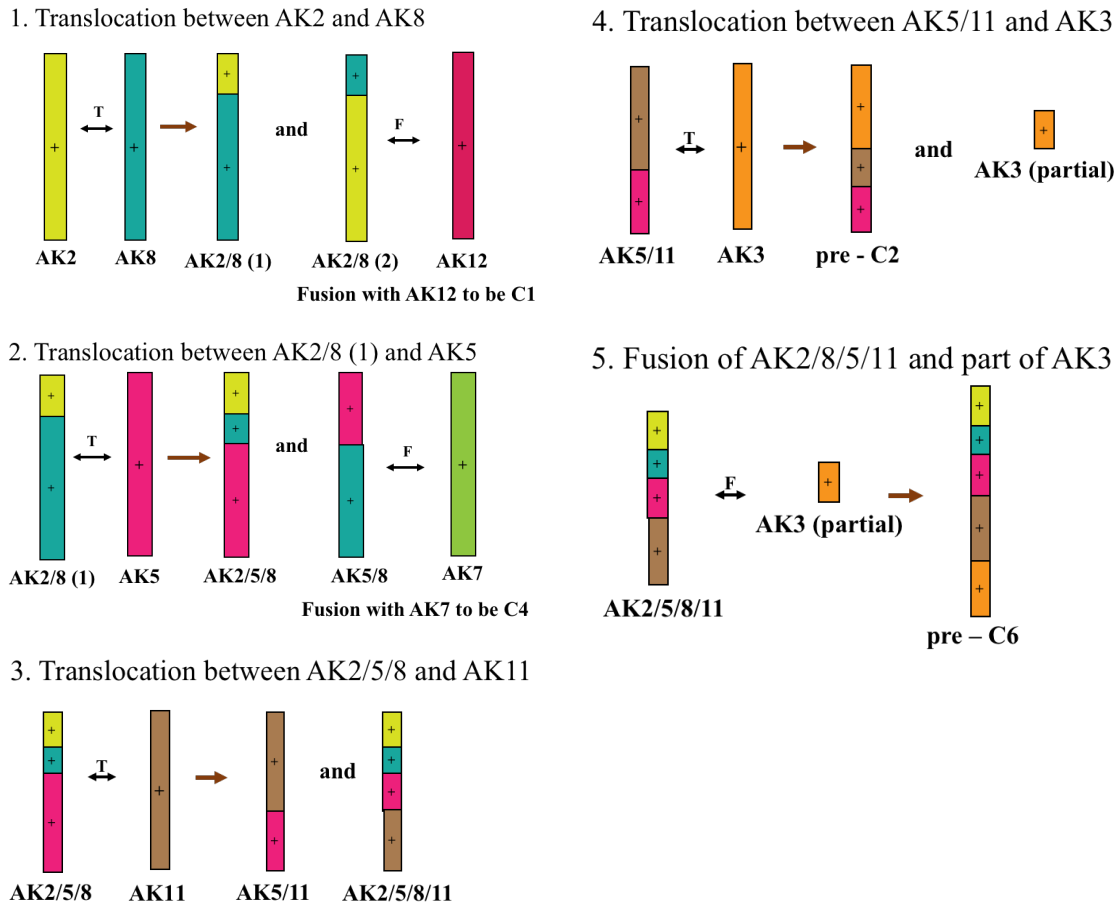


Figure 3.4 The proposed order of events for the evolution history of ancestral chromosomes AK2, AK3, AK5, AK7, AK8, AK11 and AK12 to for C1, C4, C2 and C6. T refers to reciprocal translocation and F refers to chromosome fusion.



Chapter 4 Development of an introgression library of *C. hystrix* in cucumber backgrounds

Abstract

C. hystrix (2*n* = 2*x* = 24,) is currently the only known species in the genus *Cucumis* that is sexually compatible with cucumber (*C. sativus*, 2*n* = 2*x* = 14) and has a great potential for cucumber improvement. Here, we reported the development of *C. hystrix* introgression lines (IL) in elite cucumber backgrounds through the development of an amphidiploid (2*n* = 4*x* = 38). Overall, 334 ILs have been developed and identified by SSR markers to carry *C. hystrix* alleles. The ILs included different introgression segments that covered about 57.29% of *C. hystrix* genome. The average number of introgressions was 4.8, 1, 2.6, 1, and 1 introgression segments per IL in BC_2, BC_2S_1, BC_3, BC_3S_1 and BC_4, respectively. With genotyping-by-sequencing (GBS), the introgressions in the 55 ILs included 44 syntenic blocks or about 74.57% coverage of *C. hystrix* genetic map. Each IL had an average 5.4 introgression segments and 1.7 Mbp per segment. Unilateral incompatibility and genome-wide homoeologous recombination suppression between *C. hystrix* and cucumber chromosomes were observed. Possible recombination hotspots for *C. hystrix* were identified. The low success rate of advancing BC_1 plants to next generation was the rate-limiting step in IL development for *C. hystrix*.

Introduction

One of the limitations of modern crop breeding is the reduction in available genetic variability in cultivars for most crop species, resulting from successive generations of artificial selection. Wild relatives, on the other hand, maintain a wide range of allelic diversity for important traits such as disease resistance and enhanced stress tolerance. Interspecific hybridization is used to improve crops by transferring specific traits, such as resistance to diseases or abiotic stresses, from wild relatives. Several approaches have been used to introduce DNA segments or genomes from wild species into cultivated species' genomic backgrounds. The most common approach is conventional hybridization or crossing.

In *Cucumis*, interspecific hybridization attempts have been reported. Conventional hybridization among some wild species were successful, but attempts to cross any of the wild species with the two important cultivated species in *Cucumis, C. sativus* (cucumber) and *C. melo* (melon), have failed. A cross *C. sativus* x *C. melo* resulted in fruit set but no progeny, while a reciprocal cross yielded no fruit (Deakin et al, 1971). The species *C. metuliferus, C. heptadactylus, C. humifructus, C. melo,* and *C. sativus* have never been successfully crossed with other *Cucumis* species to produce fertile F_1 hybrid plants. The following species can be crossed to a limited extent between each other: *C. africanus, C. anguria, C. dipsaceus, C. ficifolius, C. myriocarpus, C. prophetarum, C. leptodermis, C. denteri, C. sagittatus,* and *C. zeyheri* (Chen & Zhou, 2011; Deakin et al., 1971; Kho et al., 1980; Kroon et al., 1979; Singh & Yadava, 1984). Interspecific hybridization of wild species is often met with difficulties. Different species have means to remain distinct and separated from each other, such as different flowering times, different temperature suitable for growing or different genome structure. These factors lead to

interspecific crossing incompatibilities. The hybridization barriers are generally divided into two categories: pre-fertilization barrier and post-fertilization barriers.

Pre-fertilization barrier may arise due to lack of pollen germination, insufficient pollen tube growth, or pollen tube penetration of the ovule. Post-fertilization barrier arise from degeneration of a hybrid embryo, male and female sterility of a hybrid plant, and lethality in the hybrid progeny (Kaneko and Bang 2014). Post-fertilization barrier might arise due to undesirable interactions between genes of the two species interfering with cell division and differentiation, unfavorable interaction in zygotic cells between the cytoplasm and nuclear genes, unfavorable interaction between embryo, endosperm, and maternal tissue, inadequate number of fertilized ovules to prevent flower and fruit abortion, or inadequate chromosome pairing during meiosis (Fehr 1991).

Techniques to overcome pre-fertilization barrier include parental selection of compatible mating genotypes, reciprocal crosses, modification of ploidy levels to match one in the cross pair, modification of pistil, and chemical treatment of the pistil or pollen, and protoplast fusion (Fehr 1991). In *Cucumis*, protoplasts of individual species *C. sativus, C. melo, and C. metuliferus* were successfully isolated, but protoplast fusions of different species have never succeeded (Jarl et al, 1995; Roig et al, 1986; Tang & Punja, 1989). Fusion of melon and cucumber protoplasts reported development of heterokaryons into the callus stage, but further development to regenerate plants was not observed due to analogous somatic incompatibility reaction (Jarl et al, 1995).

Techniques to overcome post-fertilization barrier include parental selection of compatible mating genotypes, reciprocal crosses, use of growth-promoting substances (hormones) to delay flower abortion, use of a mixture of pollen from the female and male parents to induce cross-compatibility, embryo rescue, induction of chromosome doubling in the hybrid (Fehr 1991), creating monosomic alien chromosome addition lines (MAALs) and monosomic alien chromosome substitution lines, and use of alloplasmic lines (Kaneko and Bang 2014). Embryo culture was successfully reported for interspecific hybrids of *C. metuliferus* x *C. anguria* (Fassuliotis and Nelson 1988) and in *Cucurbita* genus (Šiško et al, 2003). Studies involving *C. hystrix* in interspecific hybridization attempts were only the synthesis of amphidiploid line through induced chromosome doubling from a *C. sativus* x *C. hystrix* F_1 interspecific hybrid (Chen et al, 2002) and the creation of two MAALs ($2n = 14 + 1$) from a cross between the cucumber and the allotriploid created from a backcross of the amphidiploid (*C. hytivus* x *C. sativus*) with pollen viability > 85% (Chen et al., 2004).

The first successful attempt of interspecific hybridization between *C. sativus* with another wild species by pollination was reported by Chen et al. (1997). A cross *C. sativus* L. x *C. hystrix* Chakr. produced fruits containing embryos with 37.3% regeneration rate. However, the F_1 hybrid, having $2n = 7 + 12 = 19$ chromosomes, was both male and female sterile. The hybrid cells could not be normally divided further via meiosis to have two daughter cells with equal genomic contents. Therefore, the synthesis of amphidiploid lines through chromosome doubling was initiated as reported in Chen et al. (2002), to which fertility was restored. The amphidiploid had $2n = 4x = 14 + 24 = 38$ chromosomes. Both homologous and homoeologous recombination

occurred normally. Nineteen bivalents were observed during cell division and resulted in two normal daughter cells with $n = 2x = 19$ chromosomes.

In this study, we aimed to develop an introgression library of *C. hystrix* in the elite cucumber Gy14 and 9930 genetic backgrounds through the development of amphidiploids and marker-assisted backcrossing, using the linkage map developed in Chapter 3 to detect the presence of *C. hystrix* introgression segments.

Materials and Methods

Plant materials

Two *C. hystrix* varieties TH1 and CN1 (accessions WI7001 and WI7002, respectively) were used as the donor genome parent. The wild accession WI7001 was originally collected from Thailand and WI7002 was collected from Southwestern China. Both accessions were maintained through cuttings and grown in the Walnut Street Greenhouse, University of Wisconsin-Madison. Recurrent parents comprised of five cucumber varieties: Gy14 (North American pickling type), WI7081A (9930 (Northern Chinese fresh market type), WI7200A (7200), WI7204A (7204), and WI7577A (G421).

IL development strategy

A *C. hystrix* F_1 plant from a cross between *C. hystrix* TH1 x CN1 was used as a female donor parent and cucumber Gy14 was used as a male recurrent parent to develop an interspecific F_1 hybrid IS1104 [$2n = 2x = 19$, HC where H stands for *C. hystrix* haploid genome ($1n = 1x = 12$) and C for cucumber haploid genome ($1n = 1x = 7$)]. Leaf and stem sections of the interspecific

hybrid IS1104 were then treated with 0.1% colchicine for 2 days to induce chromosome doubling and generate amphidiploid cells; each contained 2 full sets of chromosomes from cucumber and *C. hystrix* (*2n = 4x* = 38, HHCC). The plant sections were maintained in Murashige and Skoog (MS) medium (Murashige and Skoog 1962) in tissue culture until the seedlings could be transferred to soil. Interspecific hybrid, IS1104 and amphidiploids were developed by Dr. Yiqun Weng and Dr. Junsong Pan in 2012 (unpublished).

A selected amphidiploid, referred to as $A01S_0$ hereinafter, was backcrossed with the recurrent parents to produce five different BC_1 (allotriploid, *2n = 3x* = 26, HCC) lines. BC_1 plants were subsequently backcrossed with the recurrent parents to produce BC_2 (*2n* = 14 + 1 or 2, CC + H), BC_3 (*2n = 2x* = 14, CC), and BC_4 (*2n = 2x* = 14, CC) and, in some cases, BC_5 (*2n = 2x* = 14, CC) generation. Note that the recurrent parents were used as the pollen source at all generations. Due to difficulties in pollinating the triploid BC_1 plants, some BC_1 plants were open-pollinated in Hancock, WI field during summer with the help of bees to obtain sufficient seeds to advance to the next generation in IL development.

ILs were not necessarily backcrossed to the same recurrent parent. For instance, $A01S_0$ that was backcrossed with cucumber Gy14 recurrent parent to produce $A01Gy14BC_1$ line. Since BC_1 plants were triploid and generally sterile, it was difficult to successfully backcross with a designated or intended recurrent parent to obtain enough seeds. Thus, each BC_1 plant was backcrossed with all five recurrent parents to generate different BC_2 populations in hope that some backcrosses attempts would be successful. The IL development process was summarized

in **Figure 4.1**. *C. hystrix* F_1, the interspecific hybrid IS1104, the amphidiploid hybrid $A01S_0$, and BC_1 lines were developed by Dr. Yiqun Weng and Dr. Junsong Pan (2012, unpublished).

DNA extraction and polymerase chain reaction (PCR)

The first or second true leaf of IL seedlings were collected and vacuum-dried in a freeze-dryer. The samples then were grounded into fine powder with a high-throughput homogenizer (OPS Diagnostics, Lebanon, NJ). Genomic DNA was isolated using the CTAB method, and purified using phenol/chloroform/isoamyl alcohol (Murray and Thompson 1980). The DNA was quantified using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Tech, Wilmington, DE) and adjusted to a concentration of 50 ng/μL. PCR amplification with SSR markers and gel electrophoreses of PCR products were performed following the protocols by Li et al (2011).

Detection of C. hystrix introgression by SSR markers

The ILs were screened for introgression of *C. hystrix* segments with 96 polymorphic SSR markers selected from the *C. hystrix* genetic map developed by Yang et al. (2014), using *C. hystrix* parents TH1, CN1, amphidiploid $A01S_1$, cucumber parents Gy14, G421, 7200, 7204, and 9930 as the controls. The markers covered the 12 chromosomes of *C. hystrix*. Each marker was roughly 10-15 cM apart (**Table 4.5**). The updated locations of these SSR markers on *C. hystrix* consensus map and Gy14 v2.0 assembly was presented in **Appendix 4.1**. Starting from the BC_3 generation, plants with one or two introgression segments were self-pollinated to produce a single, homozygous introgression per line. Plants with three or more detected introgression segments were backcrossed with cucumber parents and screened with the SSR markers in the next generation. The process was repeated until only 1-2 introgressions were detected by SSR

markers, then the plants were self-pollinated for one more generation to produce a line with one homozygous introgression. The number of plants screened, and the number of plants detected to have *C. hystrix* alleles were recorded to calculate recombination frequency.

Genotyping-by-sequencing of selected ILs

One hundred and twenty ILs were sent for genotyping-by-sequencing (GBS) at the Institute of Biotechnology, Cornell University. The 120 ILs consisted of 12 BC_2S_1, one BC_2S_2, 21 BC_3, 27 BC_3S_1, one BC_3S_2, 41 BC_4, nine BC_4S_1, and seven BC_5. Pedigrees of these ILs are presented in **Appendix 4.2**. The samples were cut with *ApeK* I enzyme and the library were run on the single-ended Illumina HiSeq 2000 sequencing platform (Glaubitz et al. 2014). SNPs were called from TASSEL 5.0 pipeline (Bradbury et al. 2007) with default parameters. Both *C. hystrix* TH1 assembly and the cucumber Gy14_v2.0 assembly were used as the reference genomes in two separate SNP calls. Missing data of *C. hystrix* TH1 and CN1 and cucumber Gy14 were removed. *C. hystrix* allele introgressions were then identified from the remaining 9,029 SNPs by comparing ILs at each SNP position to *C. hystrix* and its recurrent parents.

Pollen stainability test

Pollen of cucumber parents WI7204 and 9930, the amphidiploids ($A01S_3$), BC_1, selected BC_2, BC_3, and BC_4 plants were collected and stained with 5 mM acetocarmine for 30 minutes. Pollen stainability was inspected under a light microscope (Olympus BX60) with 10x magnification. Stained pollens were counted three times for cucumber parents, and 10 times for all other materials. The percentage of stainability was calculated by counting the number of stained pollen

divided by total number of pollens counted. Two-sided, unequal variance tests were performed between generations.

Chromosome number investigation

Four ILs at BC_2S_1, BC_3, BC_4, and BC_4S_1 generations (one plant at each generation) were selected at random for chromosome counting. Root tips were harvested from vigorously growing seedlings and treated with 2 mM 8-hydroxyquinoline at room temperature for 2-4 hours in the dark and fixed in ethanol: acetic acid (3:1) solution. Chromosomes were DAPI-stained and visualized under an Olympus BX51 epifluorescence microscope. Due to difficulties in producing seeds and maintaining healthy plants, chromosome counting was not conducted for BC_2 plants. This part of work was completed by Dr. Li He at the University of Wisconsin-Madison (unpublished).

Results

Pollen stainability of ILs at different generations

Typical images of pollen staining tests are shown in **Figure 4.2**. Inviable pollens (blank, empty circles) could not be stained with acetocarmine (**Figure 4.2 G**). Pollen stainability in cucumber was high (100%) but was decreased drastically in amphidiploid and ILs at BC_1 generations. Pollen stainability then started to increase at BC_2, BC_3 and BC_4 generations (**Table 4.1**). Significant differences in pollen stainability were found between cucumber and all the other plants tested, between amphidiploid and ILs at BC_3 and BC_4 generations, and between ILs BC_1 and BC_3 and BC_4 generations ($P<0.05$) (**Table 4.2**). Pollen stainability of ILs was restored to about 80% at BC_4 generation despite the significant difference from cucumber recurrent parents.

Consistent with this, fertilization and seed production were largely comparable to cucumber at this generation. However, this experiment only included limited number of samples. There were only 7 BC_1 plants developed using different cucumber lines as the recurrent parents (Gy14, 7200, 7204, 9930, H19, and Z298), but only 5 plants were still alive at that time for pollen collection (i.e. no progeny from crosses with H19 an Z298 as the recurrent parents). BC_1 plants were triploid, and it was extremely difficult to obtain BC_2 seeds through backcrossing. While some BC_2 plants seemed to grow normally with high percentage of pollen stainability [for example, A019930BC_2-A (OP)], most BC_2 plants showed stunted growth and did not produce enough flowers for pollen collection for stainability test (no data). Other factors contributing to the low success rate of generating BC_2 plants included unsynchronized flowering time of ILs with recurrent parents, the difficulty in converting gynoecious plants to produce male flowers after silver nitrate-silver thiosulfate treatments.

Germination rate of plants at different generations

The seed germination rate was recorded at each generation during IL development. At BC_1, 4,533 fruits were harvested and 766 seeds (16.9%) were obtained. Among all fruits harvested, 181 were from hand-pollination, and the rest were from plants grown in the fields in Hancock, WI (open-pollinated, 2014-2016) and Ames, IA (9930 and Gy14 as recurrent parents, 2015-2017). BC_2 progeny were germinated in MS media with germination rate of 53.55% (211 seeds germinated from total 394 seeds), and 170 plants (80.57%) survived the seedling stage (**Table 4.3**). Germination and survival rates of ILs were improved at later generations from 34.10% to 100% germination rate and 80% to 100% survival rate in amphidiploid to BC_5 generation, respectively.

Chromosome number investigation of selected ILs at different generations

Chromosome count of PT0104 (BC_2S_1), PT0155 (BC_3), PT0188 (BC_4), and PT0222 (BC_4S_1) suggested that ILs above BC_3 generation have return to 14 chromosomes (**Figure 4.3 A-D**). A BC_2S_1 sample was included because its BC_2 parent plant appeared normal and only had one small segment introgression detected by SSR markers. Chromosome number of BC_2 plants were not investigated due to the difficulty in collecting fresh root tips from the poorly growing plants.

Overview of SSR-based IL development

Ninety-six SSRs markers were used to screen for the presence of *C. hystrix* chromatins in the ILs. The average distance between two adjacent markers was 10 cM; the shortest distance was 2.7 cM between marker 68D12 and 7G06 on chromosome 3; and the longest distance was 27.50 cM between marker 69H01 and 68E05 on Chromosome 6. Genotyping of BC_1 plants confirmed that they carried all 12 chromosomes from *C. hystrix* (data not shown).

An overview of *C. hystrix* introgressions in all 334 ILs based on 96 SSRs is shown in **Figure 4.4**. On average, there were 4.8, 1, 2.6, 1, and 1 introgressions per IL at BC_2, BC_2S_1, BC_3, BC_3S_1 and BC_4 generations, respectively (**Table 4.4**). Excluding BC_2, these ILs contained 149 *C. hystrix* segments in which 78 were different. The average number of introgression segments per chromosome at each generation varied from 1 to 3 corresponding to 0 to 55 cM in length (**Table 4.4**). Unknown introgression size referred to ILs that only had one *C. hystrix* allele at that particular position. Since SSR markers that were used for *C. hystrix* introgression screening were 10 cM apart, introgression size at one particular position varied between 0 and 10 cM.

Recombination frequency was calculated as the number of recombinant progeny divided by total number of progeny. Thus, recombination frequency for each marker was calculated by the number of ILs carrying *C. hystrix* allele divided by total number of plants screened. The recombination frequency at each marker ranged from 0 (51 markers) to 0.23 for marker 67A03 (in *C. hystrix* Chromosome 4 at 121.10 cM) (**Appendix 4.3**). Forty-five out of 96 markers showed introgression in at least one IL, accounting for 57.29% coverage of the *C. hystrix* genome. The recombination frequency varied in different chromosomes. For example, it was 0.002 (0.2%) in Chromosome 7 from 2- 52 cM, and 0.26 (26.2%) in Chromosome 10 at 37 cM. Note that recombination frequencies were calculated from combining all ILs starting from BC_3 and subsequent generations. Introgression detection at BC_2 generation was excluded. A summary of recombination frequencies by *C. hystrix* syntenic blocks and chromosomes is presented in **Table 4.5**.

Rates of introgression transfer among ILs across generations

The rate of transmission of *C. hystrix* introgressions across generations was in general low. **Table 4.6** presents three selected cases showing the rate of transfer *C. hystrix* alleles from BC_2 to BC_3 and BC_4 generations, which ranged from 0.5% to 15.9%. In all three cases, we did not observe 1:1 segregation that was expected from a backcross ($P<0.01$). In case #1, only 2 out of 39 plants (5.13%) had *C. hystrix* whole chromosome 1 introgression that was passed from the BC_2 mother to BC_3 progeny. Two out of 370 plants (0.54%) were detected to contain *C. hystrix* allele from chromosome 1 spanning 0 – 63 cM that were passed from the BC_3 mother to BC_4 progeny. The transfer rate was higher in adjacent regions where 11 plants (2.97%) contained *C. hystrix* allele from chromosome 1 spanning 75 – 98 cM. Similar results were found in Cases #2

and #3 for *C. hystrix* introgressions of chromosomes 7 and 10, except that some fragments were lost in later generations. For instance, alleles from *C. hystrix* chromosome 10 from 0 – 4 cM were lost in BC_3 generation in Case 3. In BC_4 generation, *C. hystrix* alleles were only detected at 30 cM and 37 cM, separately.

Characterization of C. hystrix introgressions with GBS

A subset of 121 ILs were sent for high-throughput genotyping. SNPs were called from TASSEL 5.0 software using *C. hystrix* TH1 assembly as the reference genome. Fifty-five ILs were shown to have at least one putative introgression. Among them, seven ILs were at BC_2S_1/S_2, 23 at BC_3 or BC_3S_1/S_2, and 25 at BC_4 or BC_4S_1/S_2 generations. The other 65 samples sent did not have introgressions due to segregation within their families.

The 55 ILs contained 295 *C. hystrix* introgression segments from 44 syntenic blocks (out of 59 syntenic blocks, Chapter 3) or about 74.57% of the *C. hystrix* genome. Each IL had an average 5.4 introgression segments and 1.7 Mbp in length per segment (**Table 4.7**). PT0279 had 23 introgression segments. PT0265 had the longest introgression, 19 Mbp in length. This introgression covered the entire *C. hystrix* chromosome 6.

Some ILs were recorded as having no introgression, such as PT0096 and PT0099. This was because the identified *C. hystrix* allele introgressions were located in the same position on the *C. hystrix* consensus map; thus, the size of the introgressions could not be estimated. An overview of introgressions in the 55 ILs is shown in **Figure 4.5 A-D**. SNPs with *C. hystrix* allele coverage were located in 915 scaffolds, 299 of which (32.7%) have been anchored to the *C. hystrix* genetic

map. The 614 *C. hystrix* TH1 scaffolds that were not anchored to the genetic map represented introgressions at unknown locations of the *C. hystrix* genome. Anchoring these scaffolds to the genetic map would result in the IL population having higher genome coverage. For example, a dCAPS (Derived Cleaved Amplified Polymorphic Sequences) marker, 155A04, from scaffold00645 (size 53 kbp) was used to genotype the *C. hystrix* F_2 population allowing anchoring scaffold00645to *C. hystrix* chromosome 6 (data not shown).

Genetic background of ILs with unknown pedigrees

Several ILs (39 out of selected 55) were developed from BC_1 plants that were backcrossed to an unknown recurrent in an open field for seed increase purpose. Principal component analysis (PCA) was performed to see their association in parentage with known *C. hystrix* or cucumber parental lines. The data used in this analysis included 55 ILs with 14,608-SNP information. A PCA plot with PC1 and PC2 is shown in **Figure 4.6 A**, which clearly shows that the genetic background of all 55 ILs was closer to that of cucumber as expected.

PCA was also conducted without *C. hystrix* and the result PCA graph is shown in **Figure 4.6 B**, which indicated that the 39 ILs with unknown pedigree were closely related to cucumber, especially 9930. This suggested that the unknown recurrent parent was most likely a cucumber. Nonetheless, the result did not eliminate a possibility that introgression identification in each IL may not come from *C. hystrix.* For instance, PT0262 had Gy14 and an unknown as one of its ancestors. PT0262 had been identified to have an introgression of *C. hystrix* chromosome 2 from 0 -9 cM. It was possible that alleles at those positions were not polymorphic between *C. hystrix*

and the unknown recurrent parent but were polymorphic to Gy14. In this case, the alleles were marked as having *C. hystrix* alleles introgressed in its genome.

Discussion

Unilateral interspecific incompatibility between C. hystrix and cucumber

Unilateral incompatibility occurs when pollinations between two species are successful in one direction but not in the other. In tomato, unilateral incompatibility is controlled by the genes *ui1.1* and *ui6.1* (Li & Chetelat, 2010a, 2015; Li et al., 2010b). In other species, this could happen when a species exhibits contrasting modes of inheritance for its nuclear, mitochondrial and chloroplast genomes. Banana and loblolly pine exhibit paternal transmission of mitochondrial DNA (Faure *et al.*, 1994; Neale and Sederoff, 1989), while Douglas-fir and kiwifruit exhibit paternal transmission of chloroplast DNA (Chat *et al.*, 1999; Neale *et al.*, 1986). Carrot exhibit paternal transmission of both plastid genomes (Boblenz et al. 1990). In *Cucumis* genus, the mitochondrial genomes of squash and watermelon were maternally transmitted, but paternally transmitted in melon and cucumber (Havey *et al.*, 1998; Wei *et al.*, 2011).

A cross between *C. hystrix* and cucumber in either directions resulted in an interspecific hybrid that was $2n = 2x = 19$ chromosomes and therefore sterile. A synthetic amphidiploid (CCHH) of an interspecific F_1 hybrid when a *C. hystrix* plant was used as a male donor was sterile. This amphidiploid had both nuclear and chloroplast genomes from cucumber maternal parent and mitochondrial genome from *C. hystrix* paternal parent. In contrast, a synthetic amphidiploid (HHCC) of an interspecific hybrid when a *C. hystrix* plant was used as a female donor was partially fertile (Pan, *unpublished*). This amphidiploid had both nuclear and chloroplast genomes

from *C. hystrix* maternal parent and mitochondrial genome from cucumber paternal parent. A study by Dr. Jia Shen and colleagues (2013) confirmed that synthetic allotetraploids derived by interspecific hybridization between *C. hystrix* and cucumber exhibit paternal inheritance of mitochondrial genome as well.

Recurrent backcrossing at BC_1 generation echoed the concept of unilateral compatibility between *C. hystrix* and cucumber. The result from **Table 4.8** showed that pollination was more successful when using BC_1 plants as a mother and cucumber as a pollen source than when crossing in the reciprocal direction. Fruit setting was significantly higher when recurrent parents were used as a pollen source (161 fruits) than when recurrent parents were used as a female (20 fruits) with $P<0.05$ (**Table 4.8**). Although the reciprocal cross yielded more seeds per fruits, seed germination was only 54% and only 42% of plants survived passing the seedling stage as compared to 80% germination rate and 60% survival rate when recurrent parents were used as a pollen source. Similar results were observed in blueberry when crossing *Vaccinium corymbosum* ($2n = 4x = 48$) with *V. virgatum* ($2n = 6x = 60$) (Miyashita et al. 2012). Due to incompatibility issues, development of introgression lines could only be done in one direction – that is, using an amphidiploid and its subsequent progeny as female parents and cucumber as a pollen source.

Limitations during the development of C. hystrix introgression lines

Developing introgression lines from *C. hystrix* through the development of amphidiploids and recurrent backcross to cucumber came with limitations. Crossing an amphidiploid ($2n = 4x = 38$) to cucumber Gy14 ($2n = 2x = 14$) resulted in triploid BC_1 ($2n = 3x = 26$). A triploid is generally sterile. At the reproductive stage, homologous pairing and recombination is a basis for

reductional cell division in meiosis I (Zamariola *et al.*, 2014). Failing to meet this condition could lead to meiotic arrest. Homoeologous pairing can still occur at a low frequency and leads to completion of normal meiotic division (Kimber and Alonso 1981). In case of *C. hystrix*-cucumber BC_1 where each cell contains 2 sets of cucumber chromosomes and 1 set of *C. hystrix* chromosomes, cucumber chromosomes will mostly pair with each other while *C. hystrix* chromosomes rarely pair with cucumber chromosomes. Thus, abnormal-shaped female flowers (**Figure 4.7 A and B**) were observed. Pollen stainability of BC_1 plants was also low, ranging from 8.5 – 30% (**Table 4.1**).

Moreover, the success rate from fertilization to germination to advance from BC_1 to BC_2 generation was quite low. This rate was calculated by (% seeds obtain) x (% germ) x (% survive) = 7.29%. However, this calculation did not account for fertility rate of BC_2 plants. Some BC_2 plants grew to adult stage but failed to fertilize successfully. Therefore, it can be concluded that advancing from BC_1 to later generations was the rate-limiting step in developing *C. hystrix* introgression lines.

Genome-wide recombination suppression between C. hystrix and cucumber

Observed recombination frequency equals to the rate of allele transfer from the parent generation to the progeny. From **Table 4.6**, *C. hystrix* introgressions were transferred from BC_2 to BC_4 generations at a low frequency. At BC_2 generation, several plants grew abnormally such as having closed female flowers that prevented pollination (**Figure 4.7 A-B**) and weird-shaped plants with stunted growth (**Figure 4.7 C-D)**. Signs of this abnormal growth suggested that some BC_2 plants might be aneuploids or whole-chromosome substitution lines. Thus, normal gametes

could not be produced due to chromosomes unable to pair, leading to meiotic cell arrest. Unfortunately, we were not able to investigate the chromosome number of BC_2 plants to confirm this hypothesis due to germination, growth and root regeneration issues.

Other explanations include interference and genome-wide recombination suppression. Plants at BC_3 and BC_4 generations appeared to grow normally and chromosome count of selected lines confirmed that ILs starting from BC_3 generation had 14 chromosomes (**Figure 4.3**). However, we did not observe a 1:1 segregation ratio in all cases ($P<0.01$). Rather, we observed higher recombination frequencies in one region than adjacent regions within the same chromosomes (**Appendix 4.3**). This result suggested interference, where a crossover in one region affected crossovers in adjacent regions.

Besides, we observed recombination at each SSR marker position at an extremely low frequency (**Appendix 4.3**) and far from the expected 1:1 segregation ratio (expected recombination rate at 0.5). The highest frequency observed was 0.23 from marker 67A03 from at the end of *C. hystrix* chromosome 4. The average frequency across all markers was 0.03. A deviation from an expected segregation ratio and an extremely low recombination frequency suggested that recombination between *C. hystrix* and cucumber was reduced genome-wide. This observation was supported in tomato species where segregation distortions were common in interspecific crosses of tomato and other species (Zamir and Tadmor 2002).

A reduction in observed recombination frequency across the *C. hystrix* genome could be due to physical constraint. When *C. hystrix* introgressions were detected from the SNP dataset using the

Gy14_v2.0 assembly as the reference, it was clear that introgressions were mostly identified on either chromosome ends and that introgressions were rarely detected in centromeric and highly heterochromatic regions along the cucumber chromosomes (**Figure 4.8 A-B**). For instance, there was no introgression detected around the centromere and heterochromatic region spanning from 20 – 27 Mbp on cucumber chromosome 2 (**Figure 4.8 A**). The same applied to cucumber chromosome 4. No introgression was identified around the centromere and heterochromatic region spanning from 12 – 17 Mbp (**Figure 4.8 B**). The centromere of cucumber chromosome 3 was only 378 kbp in size and heterochromatic regions were not defined. Fifty-one *C. hystrix* introgressions were detected in this region (Yang et al., 2012; Xu *et al.*, unpublished).

Chromosome pairing and recombination is usually suppressed around centromeres and heterochromatic regions (Lambing *et al.*, 2017). **Figure 4.9** illustrated recombination frequency of *C. hystrix* syntenic blocks along cucumber chromosomes from IL-SSR dataset. A student's t-test comparing recombination frequency of each syntenic block in heterochromatic and euchromatic regions confirmed that recombination frequency in euchromatic regions was significantly higher than that in heterochromatic regions ($P<0.05$). However, recombination frequencies in *C. hystrix* ILs were independent of syntenic block size ($r=0.04$) (**Figure 4.10**).

Furthermore, several studies suggested that sequence homology was one of the contributing factors to recombination suppression. In tomato, recombination was reduced by 27% of normal level in the F_1 *L. esculentum* x *S. lycopersicoides* hybrid relative to other tomato maps (Chetelat *et al.*, 2000). However, an increase in homoeologous recombination was obtained by crossing the *S. lycopersicoides* introgression lines to *L. pennellii*—a phylogenetically intermediate

species—or to *L. esculentum* lines containing single *L. pennellii* segments on the same chromosome (Canady *et al.*, 2006). The increase in homoeologous recombination rate was only in the regions overlapping with *L. pennellii.* Recombination is typically reduced during successive backcross generation between interspecific tomato populations to cultivated tomato despite a high degree of collinearity between the genomes of the all tomato species (Rick 1969). In *Brassica,* the B genome shares some homologous regions with the A and C genomes even though it has significantly diverged from the A and C genomes. It was observed that chromosomes between the A and C genomes pair more frequently than A-B or B-C pairing in an interspecific cross between *B. napus* and *B. carinata* (Panjabi et al. 2008; Mason et al. 2010; Navabi et al. 2011). *C. hystrix* is a wild relative to cucumber and diverged from the common ancestor about 4.6 million years ago. Structurally, *C. hystrix* has 24 chromosomes in its diploid stage, while cucumber has 14 chromosomes. The distribution patterns of 45S rDNA and 5S rDNA are also different despite high homology in the pattern of Type I/II and Type IV repeats (Han et al. 2008; Zhang et al. 2015; Li et al. 2016). Thus, sequence divergence between *C. hystrix* and cucumber is large enough to cause a recombination reduction between the two species.

Alternatively, large-scale genomic rearrangements could also explain recombination reduction. In hexaploid wheat, large-scale genomic rearrangements resulted in lower recombination rate of the B genome to the A or D genome than the recombination rate between the A and D genome (Akhunov et al. 2003). Hypothetically, the common ancestor of *C. hystrix* and cucumber had undergone 59 chromosomal rearrangement events during evolution to become 7 cucumber chromosomes (Chapter 3). H9 had a high synteny with C5 and had fused with H10 and

underwent 9 inversions to become C5. Introgressions of *C. hystrix* H9 were rarely identified both with SSR and SNP markers. Thus, genomic rearrangement could suppress homoeologous recombination between *C. hystrix* and cucumber.

Recombination hotspots in C. hystrix

Recombination patterns observed in ILs from the SSR data were not always consistent with the 55 ILs from the SNP data. Recombination rate was observed at 0.20 at the beginning of H4 from the SSR data, but only three introgressions from the 55-IL-SNP data were detected in this region. In contrast, recombination frequency was 0.23 at the end of H4 (**Appendix 4.3**), which a majority of introgressions were detected in this region (**Figure 4.5 B**). On the other hand, no introgression from H11 was detected in ILs based on the SSR data, while 79 introgressions were detected in the 55-IL-SNP data (**Figure 4.5 C**).

Nonetheless, we observed some consistency in recombination frequency between ILs from the SSR data and from the GBS data in four cases. The recombination frequency in IL-SSR at the end of H1 was relatively higher than other regions of H1 (**Appendix 4.3**). Markers in this region belonged to the syntenic block 1B3, which was syntenic to the long arm of C7. Several ILs from the 55-IL-GBS dataset were also identified to have introgressions in this region (**Figure 4.5 A and Figure 4.8 B**). The same applied for markers at the end of H4 (block 4B5), the end of H5 (block 5B6), and toward the beginning of H8 (block 8B4) where relatively high recombination frequencies were observed compared to other regions in H4, H5, and H8. Markers in these regions were syntenic to C3 (~25-30 Mb), the short arm of C2, and C4 (~20 Mb), respectively.

Several ILs from the 55-IL-GBS dataset were identified to have introgressions in these regions (**Figure 4.5 and Figure 4.8**).

Based on the assumption that introgressions detected in ILs were a result of homoeologous recombination between *C. hystrix* and cucumber chromosomes, we were able to identify at least four recombination hotspots in *C. hystrix* chromosomes: block 1B3 on the long arm of H1, block 4B5 on the long arm of H4, block 5B6 on the long arm of H5, and block 8B4 on the long arm of H8.

High-throughput genotyping aids in uncovering more introgression segments.

A set of 96 SSR markers used for screening ILs for *C. hystrix* introgressions were roughly 10 cM apart on each chromosome. Thus, segments of *C. hystrix* introgressions smaller than 10 cM were not detected and stayed hidden. For example, 717BC_2-1 was a BC_2 plant and PT0108 was its progeny (BC_2S_1). Both plants were screened for *C. hystrix* allele introgressions using SSR markers and shown to have only one introgression at the end of chromosome 1 (**Figure 4.11 A**). 717BC_2-1 was thus self-pollinated to produce a single-segment, homozygous IL. However, the GBS data revealed that PT0108 had small segment introgressions of *C. hystrix* alleles in chromosome 1, 2, 4, 5, 8, 10, 12 (**Figure 4.11 B**). Similar results were observed in barley. Genotyping a set of 73 ILs (S42ILs) originating from a cross between the spring barley cultivar Scarlett (*Hordeum vulgare* ssp. *vulgare*) and the wild barley accession ISR42-8 (*H. v.* ssp. *spontaneum*) with the 1563-SNP BOPA1 set resulted in an increased mapping precision of the introgressed segments (Schmalenbach *et al.*, 2011). Therefore, high-throughput phenotyping

should be used to better precisely identify *C. hystrix* introgressions rather than relying on genotyping with SSR markers.

Given that introgression segments detected by SSR markers were large and recombination occurs in these regions, larger segments detected by SSR markers could be broken up further upon advancing to next generations. Some segments may eventually be lost. A majority of existing ILs presented in this study had more than 3 introgression segments each and did not have introgressions, small enough in size, that could be passed stably from generation to generation. For these reasons, introgression of *C. hystrix* segments in early generations may not be stable. Suggestively, we need repeated backcrossing to advance existing ILs further for a few years to produce a final set of single, homozygous introgression lines. Canady *et al.* (2006) suggested that chromosomes with many introgressions, containing shorter regions of homology should have a higher probability of crossovers in the homoeologous segments. Furthermore, there are at least 200 ILs that have been developed but did not SNP genotyping data. Genotyping this set of ILs with SNP markers could potentially uncover more small segment introgressions.

Acknowledgement

I would like to thank Dr. Junsong Pan and Dr. Yiqun Weng for developing the interspecific F_1 hybrid and the amphidiploids between *C. hystrix* and cucumber, which laid a solid foundation for this project. I thank Dr. Yonghua Han, Dr. Li He and Dr. Jiming Jiang for cytogenetic work on chromosome count. I thank Kristin Haider for technical help, Dr. Douglas Senalik for providing a GBS SNP-calling pipeline, and Dr. Sarah Patterson for use of the microscope for pollen stainability test. I also thank Lorelei D'Huyvetter, an undergraduate student, and Alec Gillett, a

senior high school student from the YAP program, for their help in introgression screening work with SSR markers.

Reference

Akhunov ED, Akhunova AR, Linkiewicz AM, et al (2003) Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. PNAS 100:10836–41. doi: 10.1073/pnas.1934431100

Boblenz K, Nothnagel T, Metzlaff M (1990) Paternal inheritance of plastids in the genus *Daucus*. Mol Gen Genet 220:489–491

Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–5. doi: 10.1093/bioinformatics/btm308

Canady M a, Ji Y, Chetelat RT (2006) Homeologous recombination in *Solanum lycopersicoides* introgression lines of cultivated tomato. Genetics 174:1775–88. doi: 10.1534/genetics.106.065144

Chat J, Chalak L, Petit RJ (1999) Strict paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in intraspecific crosses of kiwifruit. Theor Appl Genet 99:314–322. doi: 10.1007/s001220051238

Chen J, Luo XD, Qian CT, et al (2004) *Cucumis* monosomic alien addition lines: morphological, cytological, and genotypic analyses. Theor Appl Genet 108:1343–1348. doi: 10.1007/s00122-003-1546-z

Chen J, Staub J, Adelberg J, et al (2002) Synthesis and preliminary characterization of a new species (amphidiploid) in *Cucumis*. Euphytica 123:315–322

Chen J, Staub JE, Tashiro Y, et al (1997) Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr. Euphytica 96:413–419. doi: 10.1023/A:1003017702385

Chen J, Zhou XH (2011) Wild crop relatives: Genomic and breeding resources (Vegetables). In: Chittaranjan K (ed) Wild Crop Relatives: Genomic and Breeding Resources Vegetables. Springer, pp 67–87

Chetelat RT, Meglic V, Cisneros P (2000) A genetic map of tomato based on BC_1 *Lycopersicon esculentum* x *Solanum lycopersicoides* reveals overall synteny but suppressed recombination between these homeologous genomes. Genetics 154:857–867

Deakin JR, Bohn GW, Whitaker TW (1971) Interspecific hybridization in *Cucumis*. Econ Bot 25:195–211. doi: 10.1007/BF02860080

Fassuliotis G, Nelson B V. (1988) Interspecific hybrids of *Cucumis metuliferus* x *C. anguria* obtained through embryo culture and somatic embryogenesis. Euphytica 37:53–60. doi:

10.1007/BF00037223

Faure S, Noyer JL, Carreel F, et al (1994) Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). Curr Genet 25:265–269. doi: 10.1007/BF00357172

Fehr W (1991) Principles of Cultivar Development: Theory and Technique. Macmillian Publishing Company

Glaubitz JC, Casstevens TM, Lu F, et al (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346. doi: 10.1371/journal.pone.0090346

Han YH, Zhang ZH, Liu JH, et al (2008) Distribution of the tandem repeat sequences and karyotyping in cucumber (*Cucumis sativu*s L.) by fluorescence in situ hybridization. Cytogenet Genome Res 122:80–8. doi: 10.1159/000151320

Havey MJ, McCreight JD, Rodes B, Taurick G (1998) Differential transmission of the *Cucumis* organellar genomes. Theor Appl Genet 97:122–128

Jarl CI, Bokelmann GS, De Haas JM (1995) Protoplast regeneration and fusion in *Cucumis*: melon × cucumber. Plant Cell Tissue Organ Cult 43:259–265. doi: 10.1007/BF00039953

Kaneko Y, Bang SW (2014) Interspecific and intergeneric hybridization and chromosomal engineering of Brassicaceae crops. Breed Sci 64:14–22. doi: 10.1270/jsbbs.64.14

Kho YO, Den Nijs APM, Franken J (1980) Interspecific hybridization in *Cucumis* L. II. The crossability of species, an investigation of in vivo pollen tube growth and seed set. Euphytica 29:661–672. doi: 10.1007/BF00023214

Kimber G, Alonso LC (1981) The analysis of meiosis in hybrids. II. Triploid hybrids. Can J Genet Cytol 23:221–234. doi: 10.1139/g81-026

Kroon GH, Custers JBM, Kho YO, et al (1979) Interspecific hybridization in *Cucumis* (L.). I. Need for genetic variation, biosystematic relations and possibilities to overcome crossability barriers. Euphytica 28:723–728

Lambing C, Chris F, Franklin H, Wang C-JR (2017) Update on meiotic recombination in plants: understanding and manipulating meiotic recombination in plants. Plant Physiol 173:1530–1542. doi: 10.1104/pp.16.01530

Li K, Wang H, Wang J, et al (2016) Divergence between *C. melo* and African *Cucumis* species identified by chromosome painting and rDNA distribution pattern. Cytogenet Genome Res. doi: 10.1159/000453520

Li W, Chetelat RT (2010) A pollen factor linking inter-and intraspecific pollen rejection in tomato. Science (80-) 330:1827–1830. doi: 10.1126/science.1197908

Li W, Chetelat RT (2015) Unilateral incompatibility gene *ui1.1* encodes an S-locus F-box protein expressed in pollen of *Solanum* species. PNAS 112:4417–4422. doi: 10.1073/pnas.1423301112

Li W, Royer S, Chetelat RT (2010) Fine mapping of *ui6.1*, a gametophytic factor controlling pollen-side unilateral incompatibility in interspecific *Solanum* hybrids. Genetics 185:1069–1080. doi: 10.1534/genetics.110.116343

Li Y, Yang L, Pathak M, et al (2011) Fine genetic mapping of *cp*: a recessive gene for compact (dwarf) plant architecture in cucumber, *Cucumis sativus* L. Theor Appl Genet 123:973–83. doi: 10.1007/s00122-011-1640-6

Mason AS, Huteau V, Eber F, et al (2010) Genome structure affects the rate of autosyndesis and allosyndesis in AABC, BBAC and CCAB *Brassica* interspecific hybrids. Chromosom Res 18:655–666. doi: 10.1007/s10577-010-9140-0

Miyashita C, Mii M, Aung T, Ogiwara I (2012) Effect of cross direction and cultivars on crossability of interspecific hybridization between *Vaccinium corymbosum* and *Vaccinium virgatum*. Sci Hortic (Amsterdam) 142:1–6. doi: 10.1016/j.scienta.2012.04.015

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant 15:473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Murray MG, Thompson WF (1980) Nucleic acids research rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4326. doi: doi:10.1093/nar/8.19.4321

Navabi ZK, Stead KE, Pires JC, et al (2011) Analysis of B-genome chromosome introgression in interspecific hybrids of *Brassica napus* x *B. carinata*. Genetics 187:659–73. doi: 10.1534/genetics.110.124925

Neale DB, Sederoff RR (1989) Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. Theor Appl Genet 77:212–216. doi: 10.1007/BF00266189

Neale DB, Wheeler NC, Allard RW (1986) Paternal inheritance of chloroplast DNA in Douglas-fir. Can J For Res 16:1152–1154. doi: https://doi.org/10.1139/x86-205

Panjabi P, Jagannath A, Bisht NC, et al (2008) Comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using intron polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C Brassica genomes. BMC Genomics 9:113. doi: 10.1186/1471-2164-9-113

Rick CM (1969) Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicon esculentum*: Segregation and Recombination. Genetics 62:753–768

Roig LA, Roche MV, Orts MC, et al (1986) Isolation and culture of protoplasts from *Cucumis*

metuliferus and *Cucurbita martinezii* and a method for their fusion with *Cucumis melo* protoplasts. Cucurbit Genet Coop Rep 9:70–73

Schmalenbach I, March TJ, Bringezu T, et al (2011) High-resolution genotyping of wild barley introgression lines and fine-mapping of the threshability locus *thresh-1* using the illumina GoldenGate assay. G3 1:187–96. doi: 10.1534/g3.111.000182

Shen J, Kere MG, Chen J (2013) Mitochondrial genome is paternally inherited in *Cucumis* allotetraploid (*C.*×*hytivus*) derived by interspecific hybridization. Sci Hortic (Amsterdam) 155:39–42. doi: 10.1016/j.scienta.2013.03.009

Singh AK, Yadava KS (1984) An analysis of interspecific hybrids and phylogenetic implications in *Cucumis* (Cucurbitaceae). Plant Syst Evol 147:237–252

Šiško M, Ivančič A, Bohanec B (2003) Genome size analysis in the genus *Cucurbita* and its use for determination of interspecific hybrids obtained using the embryo-rescue technique. Plant Sci 165:663–669. doi: 10.1016/S0168-9452(03)00256-5

Tang FA, Punja ZK (1989) Isolation and culture of protoplasts of *Cucumis sativus* and *Cucumis metuliferus* and methods for their fusion. Cucurbit Genet Coop Rep 12:29–34

Wei Y, Zhao G, Yang H, Chen J (2011) Inheritance analysis of mitochondrial (mt) DNA in the interspecific crossing of genus *Cucumis*. J Plant Genet Resour 12:612–618

Yang L, Koo D-H, Li Y, et al (2012) Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. Plant J 71:895–906. doi: 10.1111/j.1365-313X.2012.05017.x

Yang L, Koo DH, Li D, et al (2014) Next-generation sequencing, FISH mapping and synteny-based modeling reveal mechanisms of decreasing dysploidy in *Cucumis*. Plant J 77:16–30. doi: 10.1111/tpj.12355

Zamariola L, Tiang CL, De Storme N, et al (2014) Chromosome segregation in plant meiosis. Front Plant Sci 5:279. doi: 10.3389/fpls.2014.00279

Zamir D, Tadmor Y (2002) Unequal segregation of nuclear genes in plants. Bot Gaz 147:355–358. doi: 10.1086/337602

Zhang Y, Cheng C, Li J, et al (2015) Chromosomal structures and repetitive sequences divergence in *Cucumis* species revealed by comparative cytogenetic mapping. BMC Genomics 16:730. doi: 10.1186/s12864-015-1877-6

Table 4.1 Average and standard deviation of pollen stainability (%) of selected cucumber, amphidiploid, and introgression lines across different generations.

#	**Pollen source**	**Generation**	**Average**	**s.d.**
1	7204	cucumber parent	100.00	0.00
2	9930	cucumber parent	100.00	0.00
3	A01S_3	amphidiploid	11.19	8.09
4	A06S_1	amphidiploid	26.81	18.77
5	A09S_1	amphidiploid	45.47	26.92
6	A10S_1	amphidiploid	46.00	28.81
7	A017200RBC_1	BC_1	30.02	16.88
8	A017204BC_1	BC_1	8.51	6.52
9	A019930BC_1	BC_1	10.96	4.55
10	A019930RBC_1	BC_1	23.82	10.68
11	A25GY14BC_1	BC_1	9.83	8.34
12	A012R.3BC_2 (2)	BC_2	0.00	0.00
13	A017200RBC_2-5	BC_2	44.87	10.19
14	A019930BC_2 OP a	BC_2	79.08	11.31
15	PT0118	BC_3	68.64	21.15
16	PT0155-32	BC_3	85.14	12.83
17	PT0155-33	BC_3	70.40	19.58
18	PT0155-34	BC_3	79.01	14.31
19	PT0199-2e	BC_4	84.46	10.87
20	PT0199-2f	BC_4	71.78	18.92
21	PT0199-2g	BC_4	81.54	12.88

Table 4.2 Pairwise t-test p-values of pollen stainability of selected plants at different generations.

	cucumber	amphidiploid	BC_1	BC_2	BC_3	BC_4
cucumber	-	0.00	0.00	0.12	0.01	0.03
amphidiploid		-	0.16	0.74	0.01	0.01
BC_1			-	0.39	0.00	0.00
BC_2				-	0.27	0.24
BC_3					-	0.55
BC_4						-

Table 4.3 Seed germination record of cucumber, amphidiploid, and ILs at different generations

Generation of seedlings	**Sum of # seeds sown**	**Sum of # seeds germ**	**Sum of # dead**	**Sum of # survivals**	**% germ**	**% dead**	**% survive**
Cucumber recurrent parents	107	80	0	80	74.77	0.00	100.00
C. hystrix	134	49	8	41	36.57	16.33	83.67
Amphidiploid ($2n = 4x = 38$)	261	89	18	71	34.10	20.22	79.78
BC_2	394	211	41	170	53.55	19.43	80.57
BC_3	1,766	1,371	147	1,224	77.63	10.72	89.28
BC_4	1,305	1,170	53	1,117	89.66	4.53	95.47
BC_5	135	100	1	99	74.07	1.00	99.00
BC_6	19	19	0	19	100.00	0.00	100.00

Table 4.4 Average number of introgression segments and average size (cM) per chromosome from the ILs that were identified with 96 SSR markers.

Chr	Average number of introgressions per chromosome					Average size of introgression (cM)				
	BC_2	BC_2S_1	BC_3	BC_3S_1	BC_4	BC_2	BC_2S_1	BC_3	BC_3S_1	BC_4
H01	1.9	-	1.0	-	1.0	-	-	55.1	-	11.1
H02	1.3	-	1.5	-	-	-	-	1.7	-	-
H03	1.0	-	1.0	-	-	-	-	0.0	-	-
H04	1.8	-	1.4	-	-	-	-	2.5	-	-
H05	1.4	-	1.3	-	-	-	-	unknown	-	-
H06	1.2	-	1.0	-	1.0	-	-	38.2	-	unknown
H07	1.3	-	1.6	-	1.0	-	-	1.9	-	unknown
H08	1.0	1.0	1.0	-	1.0	-	unknown	3.5	-	unknown
H09	1.3	-	1.5	-	-	-	-	11.3	-	-
H10	1.2	-	1.2	1.0	3.0	-	-	27.8	unknown	unknown
H11	1.5	-	1.0	-	-	-	-	35.6	-	-
H12	1.2	-	1.1	-	-	-	-	unknown	-	-
Total	4.8	1.0	2.6	1.0	1.0					

Table 4.5 Recombination frequency by *C. hystrix* syntenic blocks and chromosomes. Recombination frequency was calculated from SSR marker screening data combining plants at different generations starting from BC_3.

***C. hystrix* Chr**	***C. hystrix* pos (cM)**	**Syntenic block no.**	**Syntenic block orientation**	**Gy14 pos (Mb)**	**Blcok size (Mb)**	**Recomb freq by block [a]**	**Recomb freq by chr**	**Note [b]**	**Syntenic block presence in IL-GBS data**
H01	0 - 51.23	1B1	-	0 - 5.1	5.1	0.03	0.03		yes
	51.91	1B2	-	7.5 - 16.1	8.6	n.a.		het	yes
	53.40 - 109.66	1B3	+	16.8 - 23.7	6.9	0.03		het	yes
H02	0 - 53.40	2B1	-	25.1 - 32.2	7.1	0.00	0.01		yes
	54.88 - 68.40	2B2	+	24.8 - 31.6	6.8	0.02		het	yes
	68. 40 - 98.00	2B3	+	13.8 - 19.2	5.4	0.01		het	yes
	102.89 - 127.07	2B4	-	0 - 2.7	2.7	0.02			yes
H03	0 - 9.98	3B1	+	32.1 - 32.8	0.7	0.00	0.01		yes
	9.99 - 36.27	3B2	+	0.8 - 4.2	3.4	0.05		het	yes
	38.97 - 51.95	3B3	+	18.2 - 20.6	2.4	0.00		het	yes
	52.63 - 53.98	3B4	-	21.0 -21.5	0.5	n.a.			yes
	53.9 - 55.3	3B5	+	20.9 - 27.2	6.3	n.a.		het	no
	58.2 - 104.47	3B6	-	27.2 - 31.7	4.5	0.00			yes
H04	2.16 - 55.77	4B1	-	36.2 - 41.4	5.2	0.05	0.08		yes
	55.77 - 59.42	4B2	-	25.5 - 27.9	2.4	n.a.		het	yes
	57.93 - 59.42	4B3	-	19.4 - 21.6	2.2	n.a.		het	no
	60.91 - 63.07	4B4	+	28.0 -28.9	0.9	n.a.			no
	65.91 - 125.30	4B5	-	29.4 - 36.2	6.8	0.09			yes
H05	0	5B1	-	11.7 - 11.8	0.1	n.a.	0.02		no
	0.68 - 3.65	5B2	+	17.8 - 18.5	0.7	0.000			yes
	4.32 - 15.57	5B3	-	22.7 - 23.7	1	0.02			yes
	15.58 - 25.46	5B4	+	9.6 - 11.1	1.5	n.a.		het	yes

	25.45 - 28.97	5B5	+	21.7 - 22.6	0.9	n.a.		het	no
	30.32 - 97.89	5B6	-	0 - 9.2	9.2	0.02		het	yes
H06	0 - 39.97	6B1	+	0 - 6.0	6	0.06	0.06	het	yes
	40.78	6B2	+	23.5 - 25.3	1.8	n.a.		het	yes
	40.78 - 111.39	6B3	-	6.4 - 19.2	12.8	0.05		het	yes
H07	0 - 30.96	7B1	+	27.8 - 31.1	3.3	0.01	0.01		yes
	30.96	7B2	-	27.2 - 27.8	0.6	n.a.			yes
	34.92	7B3	-	22.2 - 22.5	0.3	n.a.		het	yes
	34.25 - 91.13	7B4	-	0 - 8.8	8.8	0.00		het	yes
	91.89 - 92.56	7B5	+	10.4 - 10.6	0.2	n.a.			no
H08	1.49	8B1	-	20.9 - 21.5	0.6	n.a.	0.03	het	yes
	1.49	8B2	+	22.9 - 26.8	3.9	0.09		het	no
	2.16 - 6.49	8B3	+	8.9 - 10.2	1.3	0.00		het	no
	7.16 - 41.93	8B4	-	18.6 - 20.9	2.3	0.04			yes
	42.61 - 58.97	8B5	+	23.7 - 25.0	1.3	0.04			yes
	58.9 - 67.08	8B6	-	23.9 - 24.8	0.9	0.00			no
H09	0 - 16.0	9B1	-	31.4 - 33.3	1.9	0.00	0.01		yes
	18.84 - 26.36	9B2	-	22.0 -24.1	2.1	n.a.			yes
	26.36 - 28.39	9B3	+	18.9 - 20.8	1.9	0.01		het	no
	27.04 - 27.71	9B4	+	14.9 - 15.0	0.1	n.a.		het	no
	27.71 - 28.39	9B5	+	21.4 - 21.8	0.4	n.a.		het	no
	28.39 - 29.74	9B6	-	16.8 - 18.5	1.7	n.a.		het	no
	31.90 - 39.07	9B7	+	25.0 - 26.0	1	n.a.			yes
	41.10 - 81.16	9B8	-	0 - 3.6	3.6	0.01			yes
H10	0 - 43.33	10B1	-	26.1 - 30.8	4.7	0.07	0.06		yes
	44.14 - 56.21	10B2	+	4.0 - 13.2	9.2	0.04		het	yes
H11	0 - 21.95	11B1	-	11.8 - 16.1	4.3	0.00	0.00		yes

	24.8 - 26.4	11B2	-	34.1 - 34.7	0.6	n.a.		het	yes
	28.45 - 43.60	11B3	+	7.6 - 15.2	7.6	0.00		het	yes
	46.30 - 46.98	11B4	+	4.7 - 4.8	0.1	n.a.		het	no
	47.65 - 51.85	11B5	+	15.5 - 18.0	2.5	n.a.		het	yes
	53.87 - 58.61	11B6	+	0 - 0.7	0.7	0.00			yes
	60.77 - 79.17	11B7	-	4.9 - 7.4	2.5	0.00			yes
H12	0 - 45.69	12B1	+	3.2 - 11.8	8.6	0.06	0.02	het	yes
	47.04 - 47.72	12B2	-	32.5 - 33.2	0.7	n.a.		het	yes
	47.7 - 66.12	12B3	-	19.7 - 23.4	3.7	0.01		het	yes
	66.12	12B4	+	2.9 - 3.0	0.1	n.a.			no

Table 4.6 *C. hystrix* allele transfer from BC_2 to BC_3 and BC_4 generations in highlighted crosses of IL development.

Case	Cross	Generation	Chr: cM	# plants screened	# plants with introgressions	Recombination freq (%)	chi.test (*P*)
1	A017200RBC$_1$ x 7204	BC_2	H1, H6, H10	2	2	-	
	[A017200RBC$_1$ x 7204] x 7204	BC_3	H1: 0 - 98	39	2	5.13	2.09E-08
	[[A017200RBC$_1$ x 7204] x 7204] x 9930	BC_4	H1: 0 - 63	370	2	0.54	1.01E-80
			H1: 75 - 98	370	11	2.97	3.71E-73
2	A017204BC$_1$ x7204	BC_2	H7: 0 - 84	5	1	-	
	[A017204B$_1$ x 7204] x 7204	BC_3	H7: 0	383	69	18.02	5.88E-36
			H7: 60	383	2	0.52	1.49E-83
			H7: 84	383	1	1.00	2.04E-84
3	A01G421BC$_1$ x 9930	BC_2	H10: 0 - 50	6	1	-	
	[A01G421BC$_1$ x 9930] x 9930	BC_3	H10: 4 - 19	41	3	7.32	4.60E-08
			H10: 30 - 50	41	5	12.20	1.29E-06
	[[A01G421BC$_1$ x 9930] x 9930] x 9930	BC_4	H10: 30	302	48	15.89	2.05E-32
			H10: 37	302	2	0.66	6.51E-66

Table 4.7 Summary of number and size (bp) of introgressions in 55 ILs based on GBS data.

#	IL	# introgressions	Size of introgressions (bp)	Average size (bp)
1	PT0094	2	648,070	324,035
2	PT0096	0	0	-
3	PT0108	1	4,321,992	4,321,992
4	PT0110	3	1,113,739	371,246
5	PT0160	17	88,030,298	5,178,253
6	PT0163	2	2,146,436	1,073,218
7	PT0173	2	437,389	218,695
8	PT0176	1	154,032	154,032
9	PT0178	4	6,342,418	1,585,605
10	PT0180	3	6,948,343	2,316,114
11	PT0196	0	0	-
12	PT0214	0	0	-
13	PT0216	0	0	-
14	PT0230	1	154,032	154,032
15	PT0251	7	3,424,397	489,200
16	PT0262	11	26,317,185	2,392,471
17	PT0263	3	5,369,321	1,789,774
18	PT0265	14	34,850,667	2,489,333
19	PT0270	2	1,620,422	810,211
20	PT0275	15	27,186,634	1,812,442
21	PT0279	23	49,215,921	2,139,823
22	PT0286	3	3,966,563	1,322,188
23	PT0289	3	693,542	231,181
24	PT0290	2	1,058,751	529,376
25	PT0299	0	0	-
26	PT0300	4	6,146,670	1,536,668
27	PT0301	4	7,098,644	1,774,661
28	PT0302	5	7,263,878	1,452,776
29	PT0303	4	1,588,065	397,016
30	PT0304	4	845,633	211,408
31	PT0305	2	831,511	415,756
32	PT0307	2	1,875,666	937,833
33	PT0309	5	5,530,790	1,106,158
34	PT0310	2	454,622	227,311

#	IL	# introgression segments	Introgression size (bp)	Average size (bp)
35	PT0312	3	1,329,957	443,319
36	PT0314	2	986,575	493,288
37	PT0315	5	3,226,831	645,366
38	PT0316	3	1,514,376	504,792
39	PT0317	4	1,700,378	425,095
40	PT0318	1	169,089	169,089
41	PT0319	2	435,434	217,717
42	PT0320	7	9,274,584	1,324,941
43	PT0322	3	3,140,672	1,046,891
44	PT0326	15	25,706,964	1,713,798
45	PT0329	2	323,121	161,561
46	PT0331	14	22,842,925	1,631,638
47	PT0332	20	28,444,568	1,422,228
48	PT0333	2	531,832	265,916
49	PT0334	1	169,089	169,089
50	PT0336	18	32,143,207	1,785,734
51	PT0337	2	412,140	206,070
52	PT0339	15	28,827,764	1,921,851
53	PT0342	13	23,885,095	1,837,315
54	PT0343	3	5,213,375	1,737,792
55	PT0344	9	8,300,257	922,251
	Total	**295 segments**	**494,213,864 bp**	
	Average	**5.4 segments per IL**	**1,675,301 bp per segment**	

Table 4.8 Number of fruits, seeds, and germination rate from BC_1 plants with different pollination directions.

BC_1 plants	**Recurrent parents**	**# fruits**	**# seeds**	**# seeds per fruit**	**#seeds sown**	**#germ**	**#dead**	**% germ**	**% survive**
Male	Female	20	116	5.8	31	17	4	54.84	41.94
Female	Male	161	115	0.7	120	85	12	70.83	60.83

Figure 4.1 Strategy for *C. hystrix* introgression line development. Red color represents chromosomes from *C. hystrix* (*2n* = *2x* = 24) and green color represents chromosomes from *C. sativus* (*2n* = *2x* = 14).

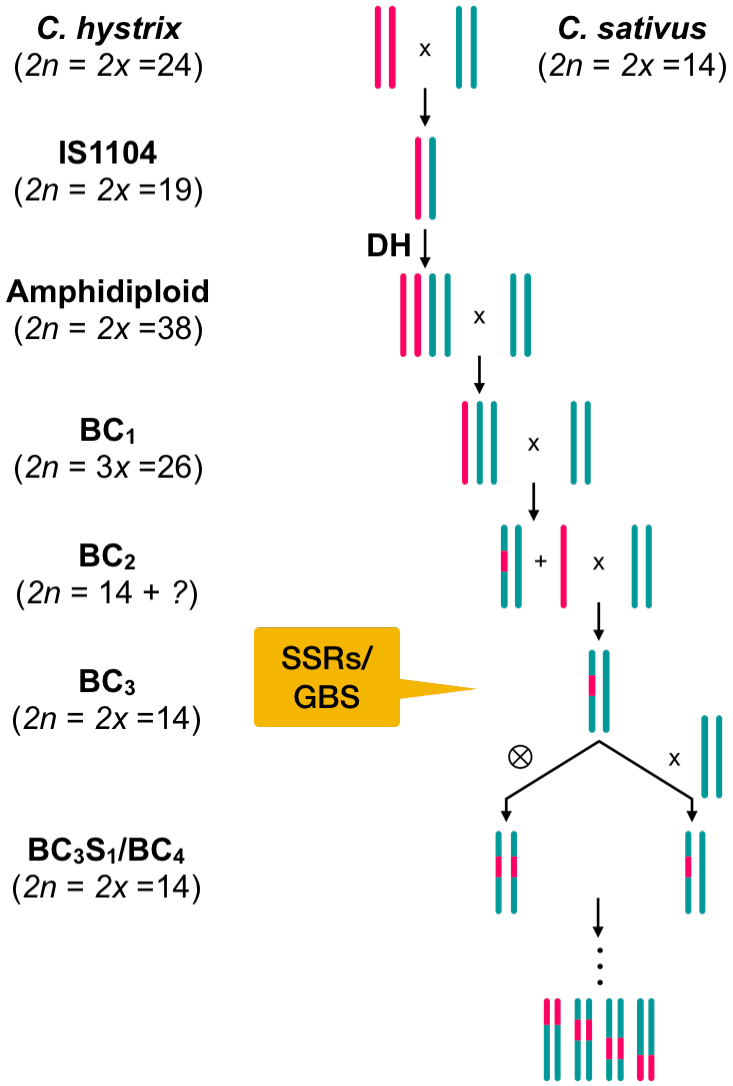


Figure 4.2 Pollen staining with 5 mM acetocarmine. A) *C. sativus*, WI7204; B) *C. sativus*, 9930; C) amphidiploid A01S_1; D) A017200RBC_1; E) A017204BC_1; F) A019930BC_1; G) A019930RBC_1; H) A25Gy14BC_1; I) A012R.3BC_2-2; J) A017200RBC_2-5; K) A019930BC_2 OP-A; L) PT0118; M) PT0155-32; N) PT0155-33; O) PT0155-34; P) PT0199-2e; Q) PT0199-2f; and R) PT0199-2g

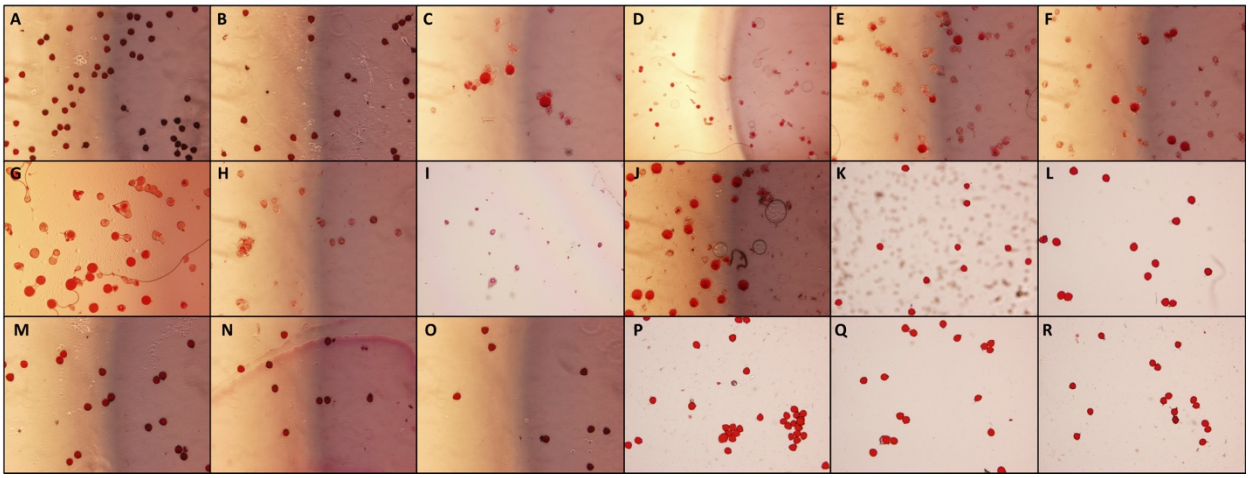


Figure 4.3 Chromosome count of A) PT0104 (BC_2S_1); B) PT0155 (BC_3); C) PT0188 (BC_4); and D) PT0222 (BC_4S_1). Bars, 10 μm. The photos were provided by Dr. Li He at the University of Wisconsin-Madison (unpublished).

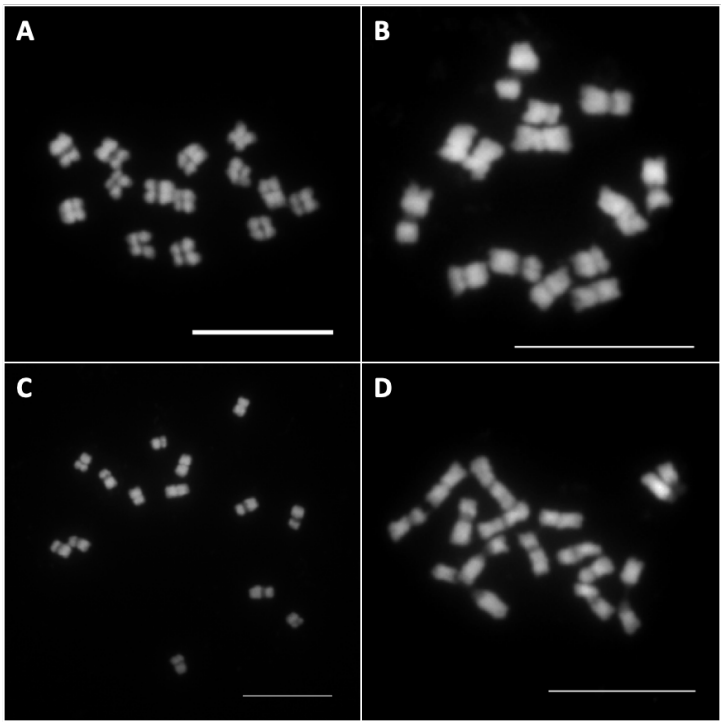


Figure 4.4 Overview of selected *C. hystrix* introgression lines (ILs) based on genotypic data from 96 SSR markers. Each row represents 1 IL. This figure includes ILs from (top) BC_2, BC_2S_1, (middle) BC_3, BC_3S_1, and (bottom) BC_4 generations, respectively.

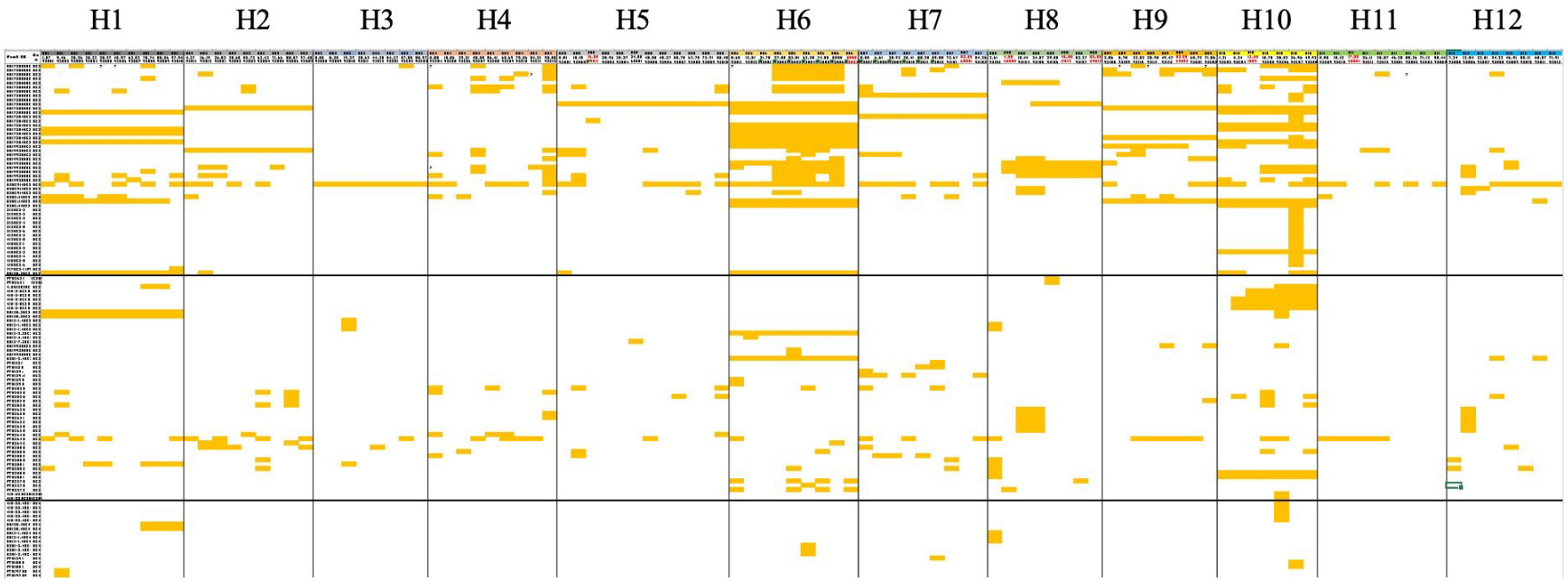


Figure 4.5 Overview of *C. hystrix* chromosome introgressions in 55 ILs based on GBS data using *C. hystrix* TH1 assembly as the reference genome in SNP calling. A) H1-H3; B) H4-H6; C) H7-H9; and D) H10-H12. A black line on each chromosome represents a centromere. Grey bar on each chromosome represents heterochromatic region. The label at the bottom of each chromosome is a distant in centimorgan (cM).

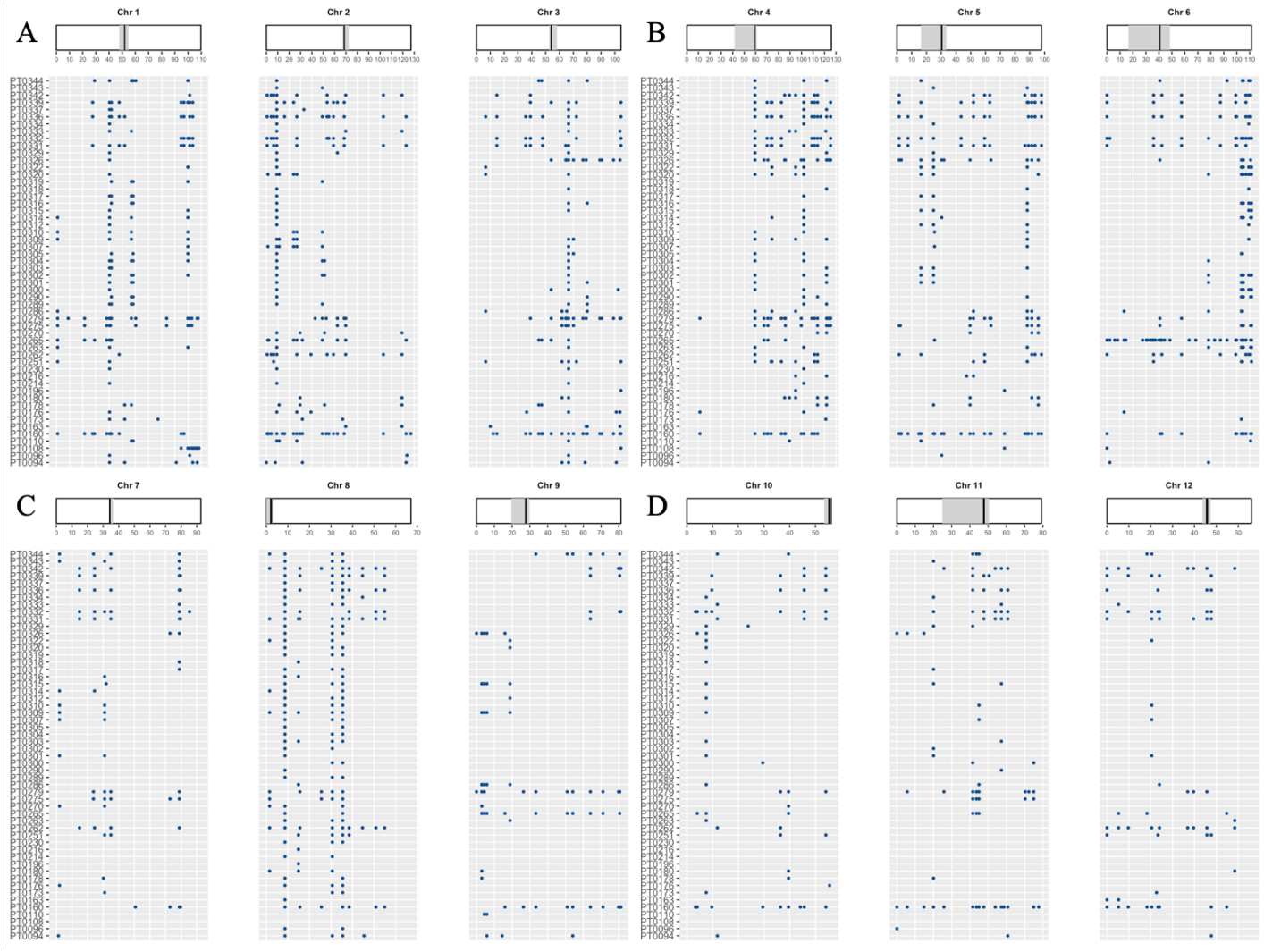


Figure 4.6 Principle component analysis for the genetic background of 55 ILs with *C. hystrix* parents and cucumber recurrent parents. A) PCA plot including *C. hystrix*; B) PCA plot excluding *C. hystrix*, "IL" represents introgression lines with known pedigree and "unknown" represents introgression lines with an unknown recurrent parent.

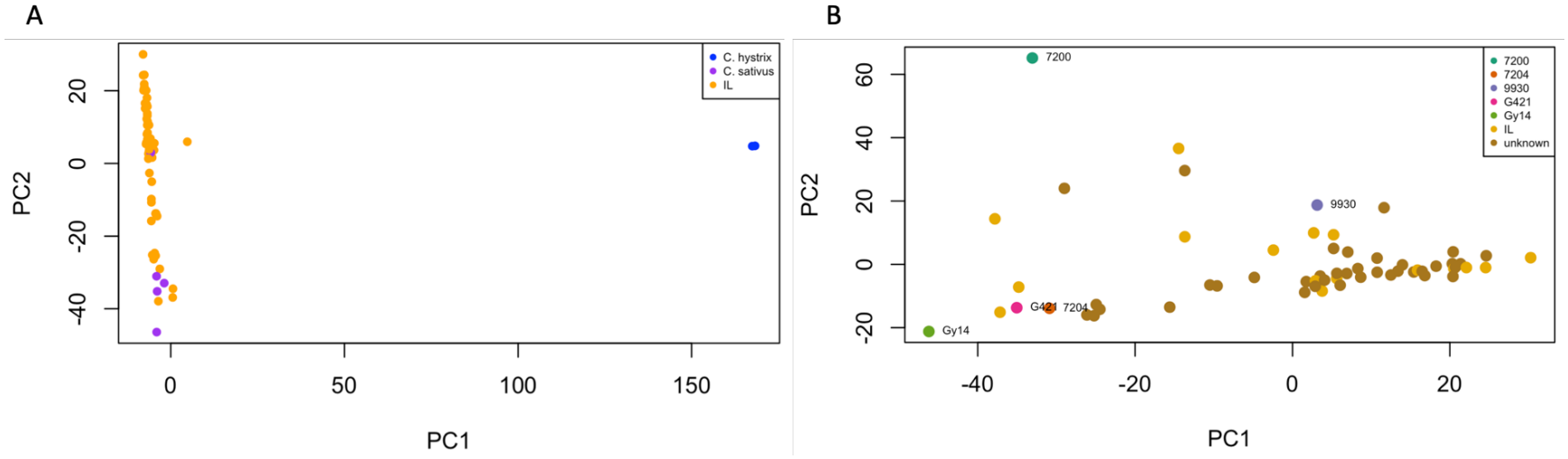


Figure 4.7 Abnormal growth of ILs at BC_2 and BC_3 generation. A) A017200RBC_2 OP-g; B) A017204BC_2-5; C) A019930BC_2-1; D) A25Gy14BC_2-2 x 9930 (PT0147)

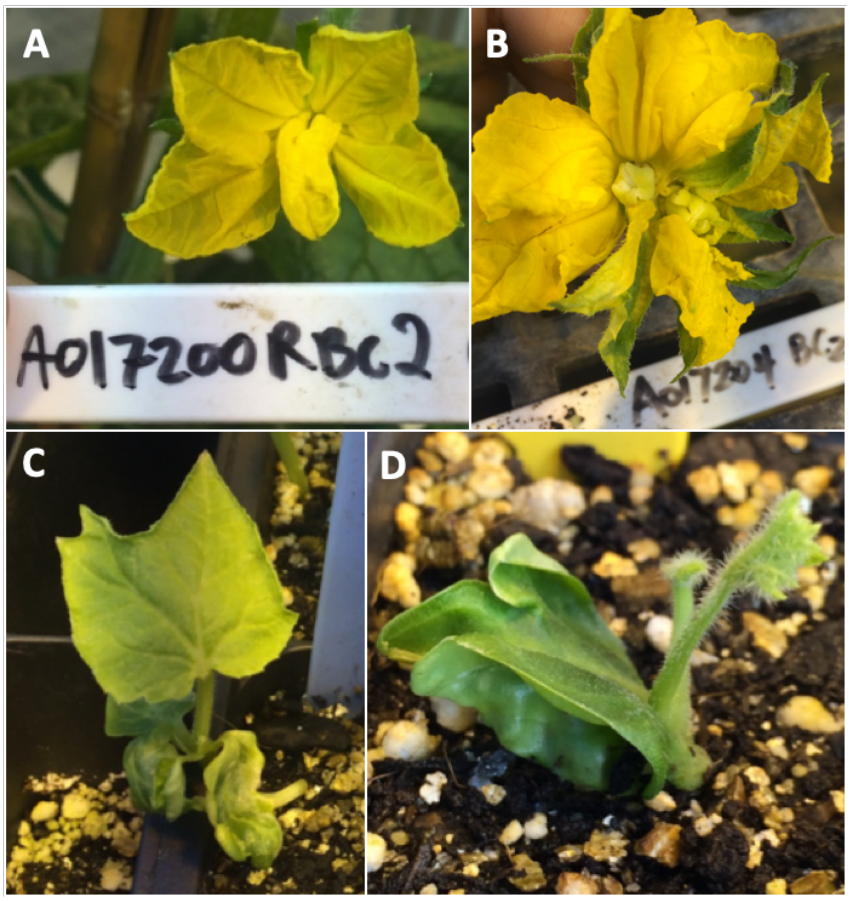


Figure 4.8 Global view of *C. hystrix* introgressions in 55 ILs based on GBS data using Gy14_v2.0 assembly as the reference genome in SNP calling. A) C1-C3 and B) C4-C7. A black bar on each chromosome represents a centromere. The label at the bottom of each chromosome is a physical distant in million base pairs (Mbp).

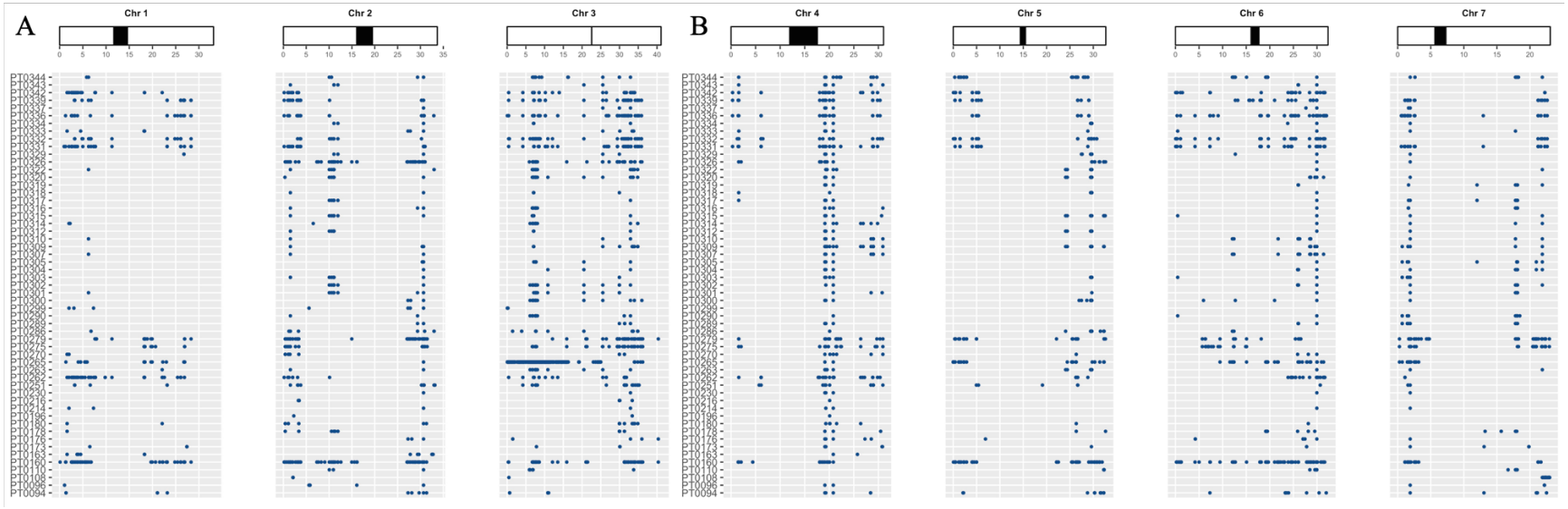


Figure 4.9 Recombination frequency of *C. hystrix* syntenic blocks along cucumber chromosomes. The frequency was calculated from all ILs beginning from BC_3 generation that were screened for introgressions using 96 SSR markers.

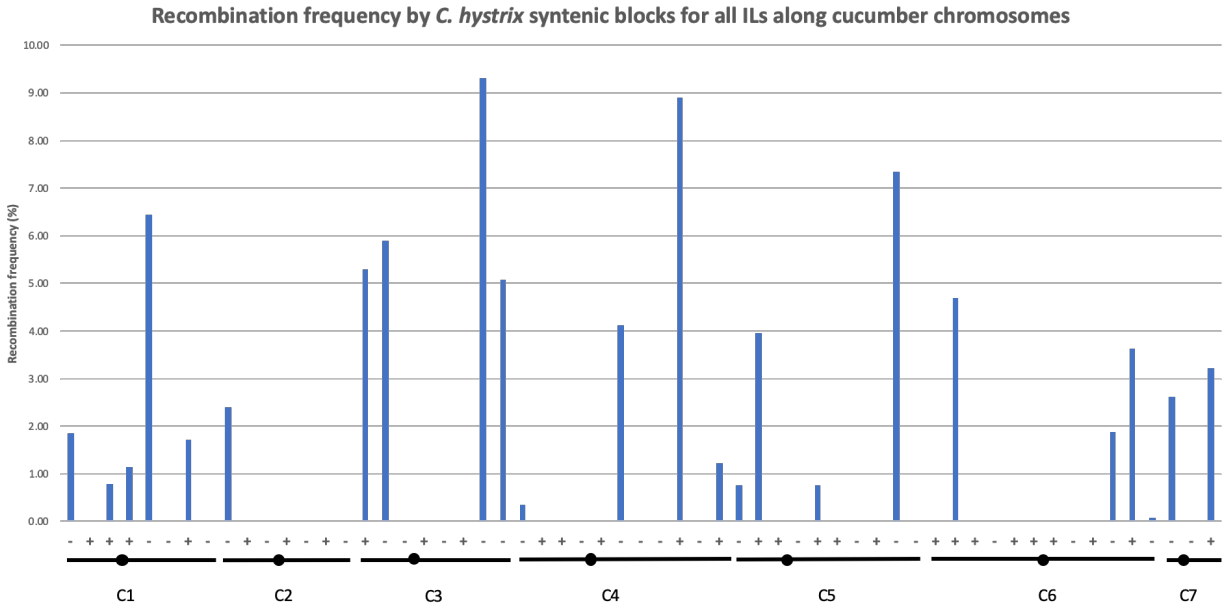


Figure 4.10 Correlation plot between the recombination frequency (%) of *C. hystrix* syntenic blocks and syntenic block size (Mb), r=0.04.

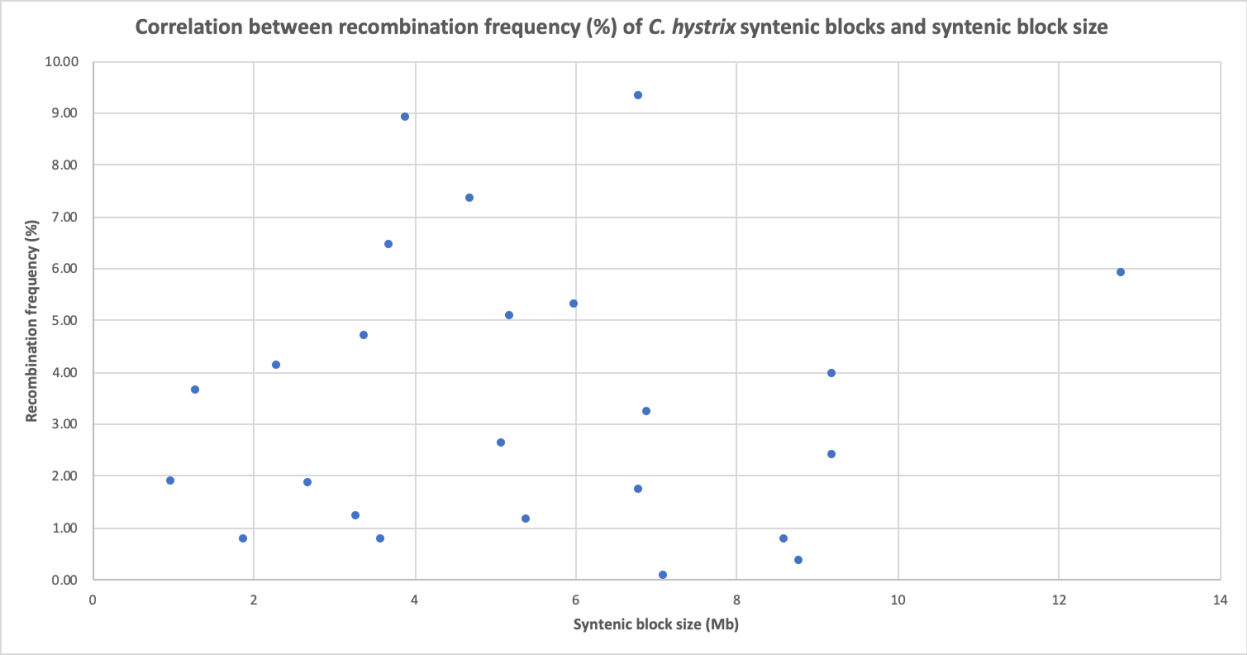
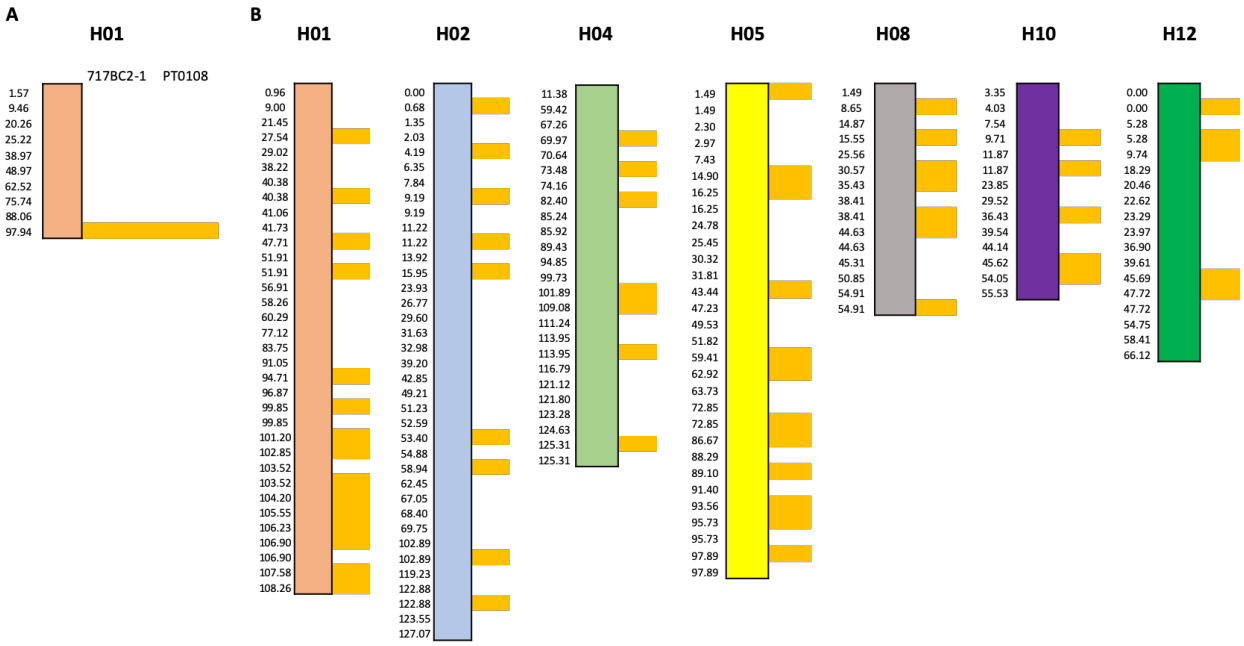


Figure 4.11 Difference in introgression identification between low-throughput genotyping (96 SSR markers) and high-throughput genotyping (GBS). A) *C. hystrix* introgressions identified by SSR markers of an introgression line PT0108 (BC_2S_1) and its maternal parent (717BC_2-1). B) *C. hystrix* introgressions identified by SNPs from GBS dataset of PT0108. Orange bars on the side of each chromosome represents *C. hystrix* introgression.



Appendix 4.1 List of 96 SSR markers that were used for *C. hystrix* introgression detection, positions on *C. hystrix* genetic map (Yang et al., 2014), and positions on Gy14 v.20 assembly.

The Appendix 4.1 is available at the following link:
https://cranberry.vcru.wisc.edu/nextcloud/index.php/s/xyAxJorLr3j9mJB

Appendix 4.2 Information about cross and pedigree for the 120 ILs that were sent for GBS at the Institute of Biotechnology, Cornell University.

The Appendix 4.2 is available at the following link:
https://cranberry.vcru.wisc.edu/nextcloud/index.php/s/xyAxJorLr3j9mJB

Appendix 4.3 The number of plants with *C. hystrix* alleles (raw data) and recombination frequency at each position detected by SSR markers at different generations. The SSR marker information was published in Yang et al, (2014).

The Appendix 4.3 is uploaded at the following link:
https://cranberry.vcru.wisc.edu/nextcloud/index.php/s/xyAxJorLr3j9mJB

Chapter 5 Preliminary screening for downy mildew and angular leaf spot resistance in selected *C. hystrix* introgression lines

Abstract

C. hystrix possesses resistance to several diseases including downy mildew (DM) and angular leaf spot (ALS). An introgression library between *C. hystrix* and cucumber has been developed. A subset of ILs were selected for phenotypic characterization of resistance to ALS and DM. Nine ILs exhibited intermediate resistance to ALS under field conditions (Hancock, WI during 2016-2018). These ILs did not have the resistance allele at the *STAYGREEN* (*CsSGR*) locus, the candidate gene for ALS resistance in the Gy14 cucumber suggesting that the source of ALS resistance in these ILs possibly is from *C. hystrix*. Four ILs exhibited high resistance to DM infection in controlled environments. Genotyping of these DM resistant ILs and several lines carrying known DM resistance QTL (*dm*1, *dm4.1* and *dm5.1*) support the *C. hystrix* origin of the DM resistance in three ILs (PT0108, PT0262, and PT0279). Possible reasons for the discrepancies in IL performance to disease infection in the fields and in controlled environment conditions are discussed. Eleven ILs selected for phenotypic evaluation in this study were still segregating for disease resistance or fruit traits, suggesting that continued recurrent backcrossing is required to develop stable ILs. These introgression lines from *C. hystrix*, provide valuable resources for cucumber improvement.

Introduction

Crop wild relatives possesses numerous valuable traits that have been exploited for breeding of cultivated crops. Traits from wild relatives that have been incorporated to cultivated crops range from pest and disease resistances, yield increase, to quality traits. Introgression lines (ILs) have been developed in many crops, such as tomato, rice, wheat, rye, melon and lettuce, among others (Ali et al., 2010; Falke et al., 2009; Jeuken & Lindhout, 2003; Perpiñá et al., 2016). More than half of genes transferred from wild relatives into cultivated crops are disease resistance genes, such as resistance to tomato yellow leaf curl virus (TYLCV) and tomato spotted wilt virus (TSWV) in tomato. Other traits transferred include yield increase gene in rice, drought tolerance genes in barley, and increased soluble solid content, fruit color, and adaptation to mechanical harvesting in tomato (Hajjar and Hodgkin 2007).

Within Cucurbitaceae, only two sets of intraspecific ILs have been developed in melon in the *inodorus* (honeydew) and *cantalupensis* (cantaloupe) backgrounds (Eduardo et al., 2005; Perpiñá et al., 2016) since attempts to cross cucumber with wild species in *Cucumis* have never succeeded (Chen & Zhou, 2011; Deakin et al., 1971; Kho et al., 1980; Kroon et al., 1979; Singh & Yadava, 1984). Melon ILs were developed from intraspecific crosses between *C. melo* subsp. *melo* and *C. melo* sup. *agrestis* since interspecific crosses between melon and wild *Cucumis* species were not viable (Deakin et al., 1971). The IL populations have possessed useful traits from wild relatives and some have already been used for various breeding purposes, such as root structure-related traits, fruit quality traits, resistance to cucumber mosaic virus (CMV), and delay climacteric fruit ripening (Eduardo et al., 2005; Perpiñá et al., 2016). Natural introgressions and

breeding of exotic genes from wild species have been reported in *Cucurbita* (Paris et al., 2006) and *Citrullus* (Zamir et al., 1984).

Resistance to downy mildew (DM), root-knot nematode (RKN), cucumber mosaic virus (CMV), and zucchini yellow mosaic virus (ZYMV) has been reported in *C. hystrix* (Chen et al. 2004; Chen and Zhou 2011). The first successful attempt to cross cucumber with a wild species was reported in 1997 from a cross between *C. sativus* x *C. hystrix* (Chen et al., 1997). Therefore, an introgression library from *C. hystrix* have been developed to incorporate exotic alleles into the genetic background of cultivated cucumber (*C. sativus*). ILs were developed by generating an amphidiploid from an F_1 hybrid (*C. hystrix* x cucumber) and recurrent backcrossing with cucumber. Marker-assisted selection (MAS) of lines with target-donor introgressions and recurrent genetic background was performed in each generation. In this study, we aimed to phenotypically characterize a set of 18 *C. hystrix* ILs for multiple traits with a focus on disease resistance to ALS and DM that was evaluated in field and controlled environments.

Materials and Methods

Plant materials

Eighteen and 9 ILs were selected for preliminary screening for downy mildew and angular leaf spot resistance, respectively (**Table 5.1**). Selection criteria included clear *C. hystrix* introgressions based on SNP-based genotypic data and seed availability. At that time, strict filtering parameters were used on IL-GBS data: maximum missing<0.9, MAF>0.1, maximum MAF=0.5, polymorphism between *C. hystrix* TH1 and CN1 and cucumber Gy14, resulting a

dataset of 297 SNPs. The selected ILs were based on the data available at the time but did not necessarily represent putative introgressions listed in Chapter 4.

Phenotypic evaluation for horticulturally important traits in selected ILs

Phenotypic data of the 18 selected ILs were collected in field trials conducted in Hancock, WI in 2016, 2017, and 2018, which were designated as WI2016, WI2017, and WI2018 respectively. All lines were grown in a randomized complete block design (RCBD) with two replications per location per year and 5 plants per replication. For each plant, phenotypic data collected in WI2016, WI2017, and WI2018 included flowering time (FT), plant height at 15th node (PH15 in cm), angular leaf spot (ALS) infection disease score, and fruit traits. Internode length was calculated by dividing plant height at 15th node by 15. For flowering time, the date of the first male and female flowers were recorded and the days to anthesis after germination was calculated as the flowering time of the plant.

Phenotypic data on fruits were collected at mature fruit stage about 30-35 days post-pollination (dpp). Five fruits were harvested for measurement for families with uniform fruit traits while 1-3 fruits were harvested from each plant for measurements for families segregating for fruit traits. Fruit traits measured included fruit number (that were used for measurement only), fruit length (FL in cm), fruit diameter (FD in cm), fruit weight (FW in kg), mature fruit skin color, and spine color. Unless the ILs were segregating for ALS and/or fruit traits, in which data were collected from individual plants, ALS scores and fruit trait data were recorded as line mean.

Preliminary screening of 18 ILs for resistance to angular leaf spot

1) Screening for resistance to ALS in a controlled environment (growth chamber)

Screening for resistance to ALS in selected ILs was performed in a growth chamber. All lines were grown in a randomized complete block design (RCBD) with two replications and 5 plants per replication. The ALS strain (*Psuedomonas syringae* pv. *lachrymans*) was provided by Dr. Alyson Thornton, Harris Moran Seed Company (Sun Prairie, WI). The inoculation followed Olczak-Woltman et al. (2009) with minor modifications. To prepare the inoculum, the pathogen was cultured on King's medium B (King et al., 1954) at 28ºC for 24 hours. The resulting colonies were suspended in 100 mM $MgCl_2$ solution and their concentration was adjusted to about 1 x 10^5 to 1 x 10^6 cells mL^{-1} using optical density measurement at 600 nm (OD 600). A drop of the surfactant, Tween 20, was added to the inoculum to ensure an even spread of the inoculum on the leaf surface. Seeds were pre-germinated in Petri-dishes and planted in PRO-MIX HP BIOFUNGICIDE + MYCHORRHIZAE Growing Medium (Premier Tech Ltd. Québec, Canada) in 4x12 plastic trays. Seedlings were grown at temperatures 27ºC/25ºC (light/dark) and 12 hours photoperiod. The seedlings were inoculated at 1-true-leaf stage by spraying the abaxial side of the cotyledons and leaves. After inoculation, the seedlings were maintained at 23ºC/20ºC (light/dark) with 16 hours photoperiod and 100% RH to promote symptom development. Gy14 and 9930 cucumber lines were included in the experiment as the resistant and susceptible controls, respectively. Disease severity was rated 7 dpi based on the portion the leaf area exhibiting water-soaked, necrotic and chlorotic lesions on a rating scale of 0-9, where 0=no damage, 1=1-10%, 2=11-20%, 3=21-30%, 4=31-40%, 5=41-50%, 6=51-60%, 7=61-70%, 8=71-80%, and 9=81-100% (or dead). Plants rated 0-2 were considered resistant (R); 3-6 intermediate (I), 7-9 susceptible (S).

2) Screening for resistance to ALS in field trials

The ILs grown in Hancock, WI field (WI2016, WI2017, and WI2018) were rated for resistance to ALS under natural infection. In each trial, disease severity was assessed weekly for three times starting at 30 days after planting (dap). The rating on a scale of 0 to 9 was based on percentage of symptomatic leaf area, where 0=no damage, 1=1-10%, 2=11-20%, 3=21-30%, 4=31-40%, 5=41-50%, 6=51-60%, 7=61-70%, 8=71-80%, and 9=81-100% (or dead).

Preliminary screening of 18 ILs for resistance to downy mildew

1) Screening for resistance to DM in controlled environment (growth chamber)

Preliminary screening for resistance to DM was performed on *C. hystrix*, cucumber recurrent parents, an interspecific F_1 hybrid, and the cucumber-*C.hystrix* amphidiploid. The result showed that *C. hystrix*, the interspecific F_1 hybrid, and the amphidiploid were resistant to DM pathogen infection. Subsequent screening in selected ILs was performed in a growth chamber in 2018 (designated as GC2018). All lines were grown in a randomized complete block design (RCBD) with two replications and 5 plants per replication. The DM pathogen (*Pseudoperonospora cubensis*) was provided by Dr. Yuhui Wang (University of Wisconsin-Madison). The pathogen was cultured and maintained on DM susceptible cucumber 9930. Seeds were pre-germinated in Petri-dishes and planted in PRO-MIX HP BIOFUNGICIDE + MYCHORRHIZAE Growing Medium (Premier Tech Ltd. Québec, Canada) in 4x12 or 6x12 plastic trays. Seedlings were grown at 27°C/25°C (day/night) and 12 hours photoperiod, and then were inoculated at 1-true-leaf stage by spraying the adaxial side of leaves with DM spore suspension at a concentration of 1×10^5 spores mL^{-1} with ~50 mL per tray. After inoculation, the seedlings were incubated in the dark for 48 hours at 20°C and 100% relative humidity (RH). After that, the seedlings were

maintained at 24ºC/18ºC (light/dark), 12 hours photoperiod, and 100% RH to promote symptom development. The WI7120B and 9930 cucumber lines were included in the experiment as the resistant and susceptible controls, respectively. Scoring of disease symptoms were conducted 7 days post inoculation (dpi) based on the portion of the leaf area exhibiting water-soaked and chlorotic lesions and followed the protocol by Wang et al (2016). The rating on a scale of 0 to 9 was based on percentage of symptomatic leaf area, where 0=no damage, 1=1-10%, 2=11-20%, 3=21-30%, 4=31-40%, 5=41-50%, 6=51-60%, 7=61-70%, 8=71-80%, and 9=81-100% (or dead) (supplemental Figure S1 in Wang et al 2016). Plants rated 0-3 were considered resistant (R), 4-6 moderately resistant (I), and 7-9 susceptible (S).

A separate experiment was conducted to test for the effect of cucumber genomic background to DM resistance in resistant ILs. Four ILs with high resistance to DM and their recurrent parents were tested for DM resistance with near-isogenic lines (NIL) that carry either *dm4.1* or *dm5.1* resistance QTL derived from the WI7120 resistance source (Wang et al. 2016) in the Gy14, 9930, and 7204 genetic backgrounds. Information of these NILs is provided in **Table 5.7**. Seeds of these NILs were provided by Dr. Yuhui Wang at University of Wisconsin, Madison. All lines were grown in a randomized complete block design (RCBD) with two replications and 4 plants per replication. 7120B was included as a resistant control.

2) Screening for resistance to DM in field trial

The selected ILs were evaluated for resistance to DM under natural inoculation of field strains at Raleigh, NC in 2018 (designated as NC2018). All lines were grown in a randomized complete block design (RCBD) with two replications and 5 plants per replication. In each trial, disease

severity was assessed weekly for each line for four times starting at 21 days after planting (dap). The rating was based on a scale of 0 to 9 following Wang et al (2016). This experiment was completed by Dr. Todd Wehner at North Carolina State University. CCMC, a North China fresh market-type cucumber that is susceptible to DM pathogens (Zhang et al. 2018), was used as the susceptible control.

Association of DM resistance in ILs with known resistance sources at dm4.1 and dm5.1 loci

ILs that were resistant to downy mildew during the preliminary screening were further tested for the association of DM resistance with known resistance sources. These ILs and relevant DM resistance sources were genotypic with 15 and 17 SSR markers in *dm4.1* and *dm5.1* regions, respectively (**Table 5.6**) (Wang et al. 2016). Genomic DNA was isolated using the CTAB method and purified using phenol/chloroform/isoamyl alcohol (Murray and Thompson 1980), quantified using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Tech, Wilmington, DE) and adjusted to a concentration of 50 ng/µL. PCR amplification with SSR markers and gel electrophoreses of PCR products followed Li et al (2011).

Screening ILs for the presence resistance allele of STAYGREEN gene (CsSGR) in the ILs

STAYGREEN gene *(CsSGR)* is the candidate gene for resistances to the downy mildew (*dm1*), angular leaf spot (*psl*) and anthracnose (*cla*) in the cucumber line Gy14, having a mutation from G (resistant) to A (susceptible) allele (Wang et al. 2019). These ILs were tested for the presence or absence of the causal SNP of the *CsSGR* gene. The SNP marker sequences for *CsSGR* were as follow: stgrF1 (GCAACAGGAACCAGCTATGACCAAGGATGGTATAATTGGCCTCA),

stgrF2 (GCAACAGGAACCAGCTATGACATGACCAAGGATGGTATAATTGGCTCCG), and stgrR (TGGCCACCACTGATATGACA).

Statistical analysis

Statistical analysis of phenotypic data was performed in R (version 3.5.1, https://www.r-project.org/). Line means and standard deviation of each phenotypic trait were calculated by rating time, replication, and environment. The Spearman's rank correlation coefficient was calculated across different rating times and environments using the grand mean of each IL. Paired student's t-tests were performed across lines in the cucumber background test.

Results

Phenotypic characterization of ALS resistance in C. hystrix ILs

Field data collection in WI2016 and WI2017 showed that *C. hystrix* TH1 and CN1 accessions and the interspecific F_1 hybrid were resistant to ALS pathogen infection (**Figure 5.1**). Thus, *C. hystrix* and the interspecific F_1 hybrid were not grown in Hancock field in 2018 and were not included in the growth chamber experiment due to slow seed germination and slow root regeneration from plant cuttings. Phenotypic data of ALS inoculation responses among 9 ILs were collected from four different environments designated as WI2016, WI2017, WI2018 and GC2018. The grand means and standard deviation of ALS disease scores for *C. hystrix* parents (TH1 and CN1), the interspecific F_1 hybrid, cucumber 9930 and Gy14 and the ILs of the last rating time across 4 environments were presented in **Table 5.2**. Cumulative mean disease scores of ILs from 30 days after planting (dap) to 58 dap across different years (under field conditions) and in the growth chamber experiments are presented in **Figure 5.2 A-D**, respectively.

Performance of different ILs to ALS resistance under field and growth chamber conditions is presented in **Figure 5.3 A-B**. Two ILs (PT0307 and PT0317) were highly resistant to ALS, one IL (PT0301) was segregating for ALS resistance, and the remaining six exhibited intermediate resistance under field conditions. However, PT0307 and PT0317 were segregating for ALS resistance in growth chamber experiments; PT0336 showed intermediate resistant to ALS, and the other six ILs were susceptible.

WI2016_rating2 was dropped from the analysis due to insufficient time for field data collection. Phenotypic correlation (*r*) of angular leaf spot disease score across different environments and rating times are presented in **Table 5.3**. Significant correlations ($P<0.05$) were found in the third and fourth ratings among field data collected from different years. Thus, ALS disease scores from the last rating of each year were used for analysis.

Preliminary screening for resistance to DM in C. hystrix

Preliminary screening for DM resistance in parental lines revealed that *C. hystrix* accessions TH1 and CN1, the interspecific F_1 hybrid (IS1104), and an amphidiploid ($A01S_1$) were completely resistant to DM pathogen infection. No symptoms of infection were observed on the C. hystrix plants or their F_1 hybrid (**Figure 5.4**). The DM resistant line 7120B showed slight chlorosis symptom of DM infection. Gy14 showed symptoms of DM infection (yellow leaf, water-soaking on the abaxial side and chlorosis), but the symptoms were less severe than those appeared on cucumber 9930, 7200, and 7204. For 7200, there were more DM spores on the abaxial side of leaves compared to Gy14, but the symptoms were not clearly visible in this picture. DM disease scores were not recorded for this test, but we concluded that *C. hystrix* and 7120B were highly

resistant to DM infection; Gy14 had intermediate resistance, and 9930, 7200, and 7204 were susceptible to DM.

Phenotypic characterization of DM resistance in C. hystrix ILs

Phenotypic data of DM inoculation responses among 18 ILs were collected from field trial (NC2018) and the controlled environment (GC2018). The grand means and standard deviation of DM disease scores for the ILs and cucumber controls across two environments and rating times are presented in **Table 5.4**. DM disease scores from the last rating under field (NC2018_rating4) and controlled environments (GC2018_rating2) are illustrated in **Figure 5.5**. Performance of different ILs to DM resistance under the controlled environment is shown in **Figure 5.6**. In general, the ILs and cucumber controls showed better DM resistance under the controlled environment than under the field conditions. Since the experiment was conducted in the growth chamber prior to field evaluation, the resistant/susceptible category assigned to each IL was based on performance in the growth chamber. Six IL (PT0108, PT0178, PT0262, PT0279, PT0307 and PT0343) were resistant to DM pathogen infection. PT0326, PT0336 and PT0342 showed intermediate resistance. PT0302, PT0332, and PT033 were susceptible to DM. The other six ILs were segregating for DM resistance (**Table 5.4**). Phenotypic correlation (r) of downy mildew disease score across different environments and rating times were presented in **Table 5.5**. Significant correlations ($P<0.05$) were found in NC2018_rating 3 across different environments and rating times. NC2018_rating4 was significantly correlated with other rating times in the field, but not with the ratings in the growth chamber.

Association of dm4.1, dm5.1, and dm1 QTL with DM resistance in the ILs

Preliminary phenotypic characterization of DM resistance showed that four ILs, PT0108, PT0262, PT0279, and PT0343 were highly resistant to DM (scores 0-1). They were tested for possible contribution by the *dm4.1* and *dm5.1* resistance loci to be observed DM resistance. The controls included were *C. hystrix* TH1 and CN1, an amphidiploid $A01S_1$, PI197088 (DM resistant), 7120B (DM resistant), and recurrent parents 9930, Gy14, 7204, and Z298 for IL development. PCR amplification results indicated that PT0108 and PT0262 did not carry the *dm4.1* and *dm5.1* resistance alleles in their genetic background. However, PT0279 had *dm4.1* and *dm5.1* in its genetic background and PT0343 had *dm4.1* in its genetic background (**Table 5.6**).

Four ILs with high resistance to DM in the preliminary test and their recurrent parents were tested for DM resistance with near-isogenic lines (NIL) that contained either *dm4.1* or *dm5.1* region in the background of Gy14, 9930, and 7204. The grand means and standard deviations (SD) of the disease scores (raw data), genetic background and downy mildew QTL region of the ILs and NILS and cucumber controls are presented in **Table 5.6**. Performance of different ILs to DM resistant compared to different near-isogenic lines (NILs) that had been introgressed with *dm4.1* or *dm5.1* in different cucumber genetic backgrounds is shown in **Figure 5.8**. Briefly, PT0108 exhibited higher resistance to DM compared to both of its recurrent parents, Gy14 and Z298, and B5-4 (*dm4.1* in Gy14 background) with $P<0.05$. PT0262 performed better than Gy14 ($P<0.05$) but similar to the resistant control 7120B and to B5-4 and E8-5, which had *dm4.1* and *dm5.1* introgressed in Gy14 genomic background. PT0279 performed better than Gy14 and 7204 ($P<0.05$), but similar to B5-4 and E8-5 (Gy14 background) that had been introgressed with

dm4.1 and *dm5.1*, respectively and to H12-16 (7204 background with *dm5.1*) with $P>0.05$. PT0343 was more resistant to DM than its recurrent parent, Gy14 and 9930, and NIL4-85 (9930 background) with $P<0.05$, but similar to B5-4 (Gy14 background), both having *dm4.1* introgressed in their genomes (**Table 5.7**).

Cucumber lines WI2757 and Gy14 had the resistance allele of the *STAYGREEN* gene (*CsSGR* or *dm1*) in their genomic backgrounds. Eight ILs, PT0108, PT0342, PT0178, PT0214, PT0303, PT0304, PT0317 and PT0333, did not have the resistance allele of *CsSGR* gene in their genomic background. The other ten ILs were heterozygous at this locus (**Figure 5.9**).

Phenotypic evaluation for horticulturally important traits in selected ILs

The grand means and standard deviation of flowering time, branch number, plant height measured up to the 15th node, internode length (in centimeter) measured up to the 15th node, total plant height, total number of nodes per plant, internode length (cm) measured from total plant height, and fruit traits including fruit skin color, spine color, length and diameter (cm), and fruit weight (in grams) for *C. hystrix*, cucumber recurrent parents, PT0108, PT0262, PT0279, and PT0343 were presented in **Appendix 5.1**. *C. hystrix* TH1 accession did not produce any flower in Hancock fields during the summer, and *C. hystrix* CN1 had extremely late flowering time (61 days MFT and 84 days FFT). PT0279 had many branches that were comparable to *C. hystrix* and the interspecific F_1 hybrid, IS1104 (8-10 branches from 2016 data); the other three ILs had branch number more comparable to cucumber than to *C. hystrix.* The internode length measured up to the 15th node of plants showed that *C. hystrix* TH1 had the shortest internode length (3.8 cm), cucumber internode length ranged from 4-6 cm and the internode length of the four ILs

resembled that of cucumber (varied from 4.8-7.8 cm). Fruits of PT0108 appeared long and Z298-like with white spine (**Figure 5.10 B**) and fruits of PT0262 appeared Gy14-like (**Figure 5.10 D**). PT0279 fruits still had black spine that was the characteristic of *C. hystrix* fruits. PT0262 was segregating for fruit traits – one group appeared closer to Gy14-like fruits with yellow skin and white spine and the other group had longer fruits but had orange skin and black spine. The picture of plants and fruits for four ILs (PT0108, PT0262, PT0279, and PT0343) were presented in **Figure 5.10 A-I**.

Discussion

The effect of rating times on evaluation of ALS resistance

Significant correlations were found between rating 3 and rating 4 under field conditions across different years (**Table 5.3**). Thus disease scores from the last rating of each year were used in the analysis. ALS evaluation was first recorded 30 days after planting, which was usually the time the first noticeable sign of ALS infection (small regions of necrosis) appeared on leaves. At this time, ALS symptoms were still not noticeable on most plants. ALS symptoms on susceptible plants or slight symptoms on resistant plants became more progressively visible weeks later (**Figure 5.2**), rendering the evaluation more accurate at a later rating time. For this reason, the last rating time was used for analyses.

Resistance to angular leaf spot in C. hystrix ILs

Discrepancies were found in phenotypic evaluations for ALS resistant in *C. hystrix* ILs between the fields and the growth chamber experiments. Most ILs showed intermediate resistance to ALS in the fields (average score 3.11 – 4.89), except for PT0307 and PT0317 that were scored as

resistant (average score 2.44), and PT0301 that was segregating (**Table 5.2**). However, all ILs were susceptible to ALS in the growth chamber tests, except for PT0336 which showed intermediate resistance. PT0307 and PT0317 were also segregating for ALS resistance in the growth chamber experiment. One explanation for this discrepancy is that different strains with lower virulence were present in the field than the one used for the growth chamber experiment. Other diseases were also present in the field, such as powdery mildew (PM). PM was also developing very early in the field. By the time ALS symptoms were first recognized, PM spores had already covered many leaves on susceptible plants. It is not known if PM infection will affect ALS symptom development. The field in Hancock, WI was also sprayed periodically with copper sulfate to limit the spread of ALS on all cucumber fields. Plant age may also be a factor. In the field, the plants were scored for ALS symptom development weekly after the ILs had been grown for 30 days in the fields, while they were evaluated at 1-true leaf stage in the growth chamber. Adult plants might have fully developed the resistance mechanism that had not been fully set up in young seedlings.

In Gy14 cucumber, ALS resistance derived from PI 197087 is controlled by a single recessive gene (*psl*), which *STAYGREEN* gene *(CsSGR)* is the candidate gene (Słomnicka et al. 2018; Wang et al. 2019). Gy14 is moderately resistant to ALS. Other recurrent parents Z298, 9930 and 7204 that were heterozygous at this locus were susceptible to ALS (**Figure 5.11**). All ILs tested for resistance to ALS either did not have the *CsSGR* resistance allele or was heterozygous at this locus. If we were to categorize ILs based on the field phenotypic evaluation, the source of the intermediate resistance to ALS in the 9 ILs should possibly come from *C. hystrix*. Alternatively, *C. hystrix* may carry the same *CsSGR* resistance allele as found in Gy14, which was not

transferred from the wild relative to those ILs during recurrent backcrossing process. A quick look at *C. hystrix* introgressions in these ILs showed that these ILs had overlapping introgression regions that belonged to the unanchored scaffolds to *C. hystrix* genetic map. They were scaffold00672, scaffold00211, scaffold00852 and scaffold10198. Unfortunately, there were no polymorphism between *C. hystrix* TH1 and CN1 for genotyping to anchor these scaffolds to the map; therefore, we could not determine which *C. hystrix* chromosome these introgressions belonged to. Last but not least, a few ILs (PT0301, PT0307, and PT0317) showed segregation of ALS resistance. Therefore, it is critical to continue repeated backcrosses of the ILs to develop ILs with stable introgressions in the genetic background.

Resistance to downy mildew in C. hystrix ILs

DM disease scores from field data collected in Raleigh, NC were generally higher than those recorded from the growth chamber experiment (**Table 5.4**). One possibility was that different races of DM were present in the field from the one used for the growth chamber experiment. The DM pathogen used for the growth chamber experiment might not be as virulent as well since the susceptible control, 9930, was only scored 6.5 and 5 in both experiments (**Table 5.4** and **Table 5.7**) on a scale of 0 – 9 where 0 was resistant and 9 was susceptible.

There are four lines with high DM resistance, which had DM disease score from 0-1 (PT0108, PT0262, PT0279, and PT0343). Detailed analyses of these lines are presented below.

1) PT0108

PT0108 did not have the resistant allele at the Cs*SGR* (or *dm1*) locus (**Figure 5.9**); neither it carried the resistance alleles at the *dm4.1* and *dm5.1* in its background (**Table 5.6**). PT0108

exhibited higher DM resistance than the two recurrent parental lines Gy14 and Z298 (P<0.05) in its pedigree (**Figure 5.7**, **Table 5.7**). It also performed well in the field with a disease score of 2-3. Thus, we concluded that *C. hystrix* introgressions contributed to resistance to downy mildew in PT0108. A quick look at *C. hystrix* introgressions in 18 ILs revealed that PT0108 had a unique introgression from H01 from 99.85 – 108.26 cM (8.4 cM), which was approximately 1.88 Mb in size. This introgression was possibly located on cucumber chromosome C7 since this region was syntenic to the long arm of C7. PT0108 was still segregating for fruit size (**Figure 5.10 B**). Repeated backcrossing was still required to develop IL(s) with stable introgressions in its genetic background. Lastly, PT0108 performed best in terms of ALS and DM resistance among all ILs and should be selected for further molecular analysis in cucumber.

2) PT0262

Resistance to DM in PT0262 did not come from *dm1* since *dm1* is recessive and PT0262 was heterozygous at this locus (**Figure 5.9**). For DM resistance, PT0262 performed better than Gy14 (P<0.05) but was comparable to 7120B (P>0.05). PT0262 also performed similar to B5-4 and E8-5, which carried *dm4.1* and *dm5.1* DM resistance QTL in Gy14 genetic background, respectively (**Table 5.7**). Genotyping with associated markers also showed that PT0262 did not have *dm4.1* and *dm5.1* in its genetic background (**Table 5.6**). However, PT0262 had Gy14 and an unknown source in its pedigree as recurrent parents. Phenotypic evaluation of other traits in the field showed that the male (MFT) and female flowering time (FFT) of PT0262 ranged from 45-48 days and 39-47 days, respectively (**Appendix 5.1**). The average MFT and FFT of WI7120 were 37 and 42 days and the average MFT and FFT of PI197088 were 41 and 42 days, respectively (Ronald Dymerski, personal communication). Flowering time of PT0262 was later

than that of the DM resistant lines WI7120 and PI197088. However, the internode length at 15th node of PT0262 (6.23 - 7.55 cm) was about twice the internode length of *C. hystrix* (3.51 - 3.80 cm). Therefore, it was possible that the source of DM resistant came from *C. hystrix* or an unknown cucumber recurrent parent that had late flowering time.

3) PT0279

Resistance to DM in PT0279 did not come from *dm1* because PT0279 was heterozygous for the allele at this locus (**Figure 5.9**). PT0279 had Gy14 and 7204 as its recurrent parents. PT0279 performed better than Gy14 and 7204 (P<0.05) in DM resistant evaluation test, but similar to B5-4 and E8-5 (Gy14 background) that had been introgressed with *dm4.1* and *dm5.1*, respectively and to H12-16 (7204 background with *dm5.1*) with P>0.05 (**Table 5.7**). It was unclear why the PCR amplification result showed that PT0279 had a similar band pattern to that of WI7120 in both *dm4.1* and *dm5.1* regions. PT0279 also had new band pattern that was not similar to any controls' band patterns (**Table 5.6**). However, previous studies have shown that rapid genome changes usually occur following allopolyploidization to stabilize the genome (Comai *et al.*, 2000; Madlung *et al.*, 2002; Shaked *et al.*, 2001). A newly synthesized allotetraploid (amphidiploid) from an interspecific F_1 hybrid of *C. hystrix* x cucumber was detected to have lost some parental restriction fragments and gained novel fragments (Chen *et al.*, 2007). New PCR band pattern and 7120B band pattern that were detected in PT0279 could possibly result from the gain of novel fragments in the amphidiploid that were passed on to the next generation during the IL development. Since PT0279 had a known pedigree, the source contributing to DM resistant in this line possibly came from *C. hystrix.*

4) PT0343

Resistance to DM in PT0343 did not come from *dm1* since this line was heterozygous at *dm1* locus (**Figure 5.9**). PT0343 had Gy14, 9930 and an unknown source in its parentage. PT0343 was more resistant to DM than Gy14 and 9930 ($P<0.05$). PT0343 also performed better than NIL4-85 (9930 background) with $P<0.05$, but similar to B5-4 (Gy14 background) with $P>0.05$ (**Table 5.7**), both having *dm4.1* introgressed in their genomes. However, genotyping revealed that PT0343 had *dm4.1* in the background (**Table 5.6**). MFT and FFT of PT0343 were also similar to that of WI7120 and PI197088. The internode length at 15th node of PT0262 (5.22 – 6.66 cm) was also longer than the internode length of *C. hystrix* (3.51 - 3.80 cm). Thus, the source of DM resistance in PT0343 might come from the unknown cucumber recurrent parent with similar resistance to cucumber 7120B. Alternatively, both *dm4.1* and the introgressed *C. hystrix* allele contributed to DM resistance in PT0343. Lastly, field evaluation showed that PT0343 was segregating for fruit traits (**Figure 5.10 H-I**). Additional backcrossing is needed for stable introgressions in the genetic background. A quick scan of PT0262 and PT0279 revealed introgressions of *C. hystrix* H12 between 36.90 – 39.61 cM (2.71 cM in size) and four unanchored scaffolds (scaffold00862, scaffold01218, scaffold01401 and scaffold02753) that were not shared with other susceptible ILs included in the DM evaluation.

In crop breeding, significant efforts are invested in developing disease resistant lines originating in wild relatives (Zamir 2001; Hajjar and Hodgkin 2007). Introgression lines are useful for mapping projects and studying the effect of individual QTL and interactions between QTLs (Gur and Zamir 2004; Muir and Moyle 2009). The ILs for ALS and DM resistances developed in this study should be valuable resources as pre-breeding materials for cucumber improvement.

Acknowledgement

I would like to thank Dr. Yuhui Wang for providing *P. cubensis* isolates and the DM-resistant near-isogenic lines and for help with DM screening, Dr. Alyson Thornton, Harris Moran Seed Company (Sun Prairie, WI) for *P.* syringae isolates and for help with ALS screening process. I thank all members of Weng lab for help with data collection in the field. I thank Ms. Kathleen Reitsma at the North Central Plant Introduction Station in Ames, IA for allowing me to use the facility to help with seed increase for my project. I thank Dr. Todd C. Wehner for DM field data collection in Raleigh, NC during summer 2018.

Reference

Ali ML, Sanchez PL, Yu S, et al (2010) Chromosome Segment Substitution Lines: A Powerful Tool for the Introgression of Valuable Genes from *Oryza* Wild Species into Cultivated Rice (*O. sativa*). Rice 3:218–234. doi: 10.1007/s12284-010-9058-3

Chen J, Moriarty G, Jahn M (2004) Some disease resistance tests in *Cucumis hystrix* and its progenies from interspecific hybridization with cucumber. In: Proceedings of Cucurbitaceae 2004, the 8th EUCARPIA Meeting on Cucurbit Genetics and Breeding. pp 189–196

Chen J, Staub JE, Tashiro Y, et al (1997) Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr. Euphytica 96:413–419. doi: 10.1023/A:1003017702385

Chen J, Zhou XH (2011) Wild Crop Relatives: Genomic and Breeding Resources (Vegetables). In: Chittaranjan K (ed) Wild Crop Relatives: Genomic and Breeding Resources Vegetables. Springer, pp 67–87

Chen L, Lou Q, Zhuang Y, et al (2007) Cytological diploidization and rapid genome changes of the newly synthesized allotetraploids *Cucumis* x hytivus. Planta 225:603–14. doi: 10.1007/s00425-006-0381-2

Comai L, Tyagi AP, Winter K, et al (2000) Phenotypic Instability and Rapid Gene Silencing in Newly Formed Arabidopsis Allotetraploids. Plant Cell 12:1551–1567. doi: 10.1105/tpc.2.9.877

Deakin JR, Bohn GW, Whitaker TW (1971) Interspecific hybridization in *Cucumis*. Econ Bot 25:195–211. doi: 10.1007/BF02860080

Eduardo I, Arús P, Monforte AJ (2005) Development of a genomic library of near isogenic lines (NILs) in melon (*Cucumis melo* L.) from the exotic accession PI161375. Theor Appl Genet 112:139–148. doi: 10.1007/s00122-005-0116-y

Falke KC, Miedaner T, Frisch M (2009) Selection strategies for the development of rye introgression libraries. Theor Appl Genet 119:595–603. doi: 10.1007/s00122-009-1069-3

Gur A, Zamir D (2004) Unused Natural Variation Can Lift Yield Barriers in Plant Breeding. PLoS Biol 2:e245. doi: 10.1371/journal.pbio.0020245

Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica 156:1–13. doi: 10.1007/s10681-007-9363-0

Jeuken MJW, Lindhout P (2003) Future perspectives of Backcross Inbred Lines for exploitation of wild germplasm : a case study on *Lactuca saligna* as a donor for quantitative resistance to lettuce downy mildew. Eucarpia Leafy Veg 69–74

Kho YO, Den Nijs APM, Franken J (1980) Interspecific hybridization in *Cucumis* L. II. The crossability of species, an investigation of in vivo pollen tube growth and seed set. Euphytica 29:661–672. doi: 10.1007/BF00023214

King EO, Ward MK, Raney DE (1954) Two simple media for the demonstration of pyocyanin and fluorescin. J Lab Clin Med 44:301–307. doi: 10.5555/URI:PII:002221435490222X

Kroon GH, Custers JBM, Kho YO, et al (1979) Interspecific hybridization in *Cucumis* (L.). I. Need for genetic variation, biosystematic relations and possibilities to overcome crossability barriers. Euphytica 28:723–728

Li Y, Yang L, Pathak M, et al (2011) Fine genetic mapping of *cp*: a recessive gene for compact (dwarf) plant architecture in cucumber, *Cucumis sativus* L. Theor Appl Genet 123:973–83. doi: 10.1007/s00122-011-1640-6

Madlung A, Masuelli RW, Watson B, et al (2002) Remodeling of DNA Methylation and Phenotypic and Transcriptional Changes in Synthetic Arabidopsis Allotetraploids. Plant Physiol 129:733–746. doi: 10.1104/pp.003095

Muir CD, Moyle LC (2009) Antagonistic epistasis for ecophysiological trait differences between *Solanum* species. New Phytol 183:789–802. doi: 10.1111/j.1469-8137.2009.02949.x

Murray MG, Thompson WF (1980) Nucleic Acids Research Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4326. doi: doi:10.1093/nar/8.19.4321

Olczak-Woltman H, Bartoszewski G, Mądry W, Niemirowicz-Szczytt K (2009) Inheritance of resistance to angular leaf spot (*Pseudomonas syringae* pv. *lachrymans*) in cucumber and identification of molecular markers linked to resistance. Plant Pathol 58:145–151. doi: 10.1111/j.1365-3059.2008.01911.x

Paris HS, Burger Y, Schaffer A (2006) Genetic variability and introgression of horticulturally valuable traits in squash and pumpkins of *Cucurbita pepo*. Isr J Plant Sci 54:223–231. doi: 10.1560/IJPS_54_3_223

Perpiñá G, Esteras C, Gibon Y, et al (2016) A new genomic library of melon introgression lines in a cantaloupe genetic background for dissecting desirable agronomical traits. BMC Plant Biol 16:1–21. doi: 10.1186/s12870-016-0842-0

Shaked H, Levy AA, Feldman M (2001) Allopolyploidy-Induced Rapid Genome Evolution in the Wheat (*Aegilops-Triticum*) Group. Plant Cell 13:1749–1759. doi: 10.1105/tpc.13.8.1749

Singh AK, Yadava KS (1984) An Analysis of Interspecific Hybrids and Phylogenetic Implications in *Cucumis* (Cucurbitaceae). Plant Syst Evol 147:237–252

Słomnicka R, Olczak-Woltman H, Korzeniewska A, et al (2018) Genetic mapping of *psl* locus and quantitative trait loci for angular leaf spot resistance in cucumber (*Cucumis sativus* L.).

Mol Breed 38:111. doi: 10.1007/s11032-018-0866-2

Wang Y, Tan J, Wu Z, et al (2019) STAYGREEN, STAY HEALTHY: a loss-of-susceptibility mutation in the *STAYGREEN* gene provides durable, broad-spectrum disease resistances for over 50 years of US cucumber production. New Phytol 221:415–430. doi: 10.1111/nph.15353

Wang Y, VandenLangenberg K, Wehner TC, et al (2016) QTL mapping for downy mildew resistance in cucumber inbred line WI7120 (PI 330628). Theor Appl Genet 129:1493–1505. doi: 10.1007/s00122-016-2719-x

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989

Zamir D, Navot N, Rudich J (1984) Enzyme Polymorphism in *Citrullus lanatus* and *C . colocynthis* in Israel and Sinai. Plant Syst Evol 146:163–170

Zhang K, Wang X, Zhu W, et al (2018) Complete resistance to powdery mildew and partial resistance to downy mildew in a *Cucumis hystrix* introgression line of cucumber were controlled by a co-localized locus. Theor Appl Genet 131:2229–2243. doi: 10.1007/s00122-018-3150-2

Table 5.1 Number of SNPs associated with having C. hystrix alleles for 18 ILs selected for preliminary screening for disease resistance to downy mildew and angular leaf spot.

Reference genome	**Gy14_v2.0 assembly**	***C. hystrix* TH1 assembly**	**DM screening**	**ALS screening in growth chamber**
ILs	**# SNPs with *C. hystrix alleles***	**# SNPs with *C. hystrix alleles***		
PT0108	27	60	yes	yes
PT0178	0	111	yes	yes
PT0214	15	18	yes	yes
PT0262	19	53	yes	no
PT0279	18	65	yes	no
PT0301	16	27	yes	yes
PT0303	30	62	yes	no
PT0304	18	23	yes	no
PT0307	15	16	yes	yes
PT0317	3	24	yes	yes
PT0326	29	39	yes	no
PT0331	36	73	yes	no
PT0332	24	81	yes	yes
PT0333	2	129	yes	no
PT0336	27	83	yes	yes
PT0339	21	22	yes	yes
PT0342	25	73	yes	no
PT0343	26	19	yes	no

Table 5.2 Mean and standard deviation (SD) of angular leaf spot disease scores from the last rating (raw data) of 9 ILs, *C. hystrix*, an interspecific F_1 hybrid, and cucumber controls across different years and environments. ALS score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.

Material	WI2016 rating4	WI2017 rating3	WI2018 rating3	GC2018 rating2	Cat [d] field	Cat [e] GC
TH1[b, c]	0.00 ± 0.00	0.00 ± 0.00	n.a.	n.a.	R	-
CN1	0.00 ± 0.00	0.00 ± 0.00	n.a.	n.a.	R	-
F_1 hybrid	0.00 ± 0.00	0.00 ± 0.00	n.a.	n.a.	R	-
9930	6.00 ± 0.00	6.00 ± 0.00	8.00 ± 0.00	8.62 ± 0.51	S	S
Gy14	3.00 ± 0.00	1.50 ± 0.52	2.00 ± 0.00	2.37 ± 1.68	R	R
PT0108	n.a.	1.50 ± 0.52	3.11 ± 1.05	8.12 ± 0.64	I	S
PT0178	4.00 ± 0.00	2.50 ± 0.52	4.25 ± 0.46	8.50 ± 0.75	I	S
PT0214	4.40 ± 0.54	2.50 ± 0.52	4.88 ± 1.05	8.37 ± 0.51	I	S
PT0301	3.00 ± 0.00	2.00 ± 0.00	4.77 ± 1.20	8.00 ± 0.53	(seg) R, S	S
PT0307	3.00 ± 0.00	2.00 ± 0.00	3.00 ± 0.00	6.62 ± 2.13	R	(seg) R, S
PT0317	4.00 ± 0.00	2.00 ± 0.00	2.44 ± 0.52	7.25 ± 2.37	R	(seg) R, S
PT0332	3.00 ± 0.00	2.00 ± 0.00	3.44 ± 0.52	8.25 ± 0.70	I	S
PT0336	3.00 ± 0.00	2.50 ± 0.52	4.11 ± 1.26	6.37 ± 1.30	I	I
PT0339	n.a.	3.50 ± 0.52	4.50 ± 0.52	8.37 ± 0.74	I	S

Notes:
[a] WI2016, WI2017, and WI2018 were conducted at the Hancock Argricultural Research Station, Hancock, WI. GC2018 was conducted in the growth chamber at the University of Wisconsin, Madison.
[b] *C. hystrix* TH1 and CN1 and an interspecific F_1 hybrid were not included in the trial and the data were not available.
[c] Disease score ranged from 0 (resistant) to 9 (susceptible).
[d] Categories was assigned based on WI2018_rating3. R was resistant, I was intermediate resistant, and S was susceptible.
[e] Categories was assigned based on GC2018_rating2. R was resistant, I was intermediate resistant, and S was susceptible.

Table 5.3 Spearman's rank correlation coefficients of ALS disease scores of ILs across different years and environments. Highlighted in grey represent significant correlations (P<0.05).

	WI2016 rating1	**WI2016 rating3**	**WI2016 rating4**	**WI2017 rating1**	**WI2017 rating2**	**WI2017 rating3**	**WI2018 rating1**	**WI2018 rating2**	**WI2018 rating3**	**GC2018 rating1**	**GC2018 rating2**
WI2016 rating1	-										
WI2016 rating3	0.74	-									
WI2016 rating4	0.82	0.67	-								
WI2017 rating1	0.14	0.12	0.46	-							
WI2017 rating2	0.54	0.62	0.80	0.27	-						
WI2017 rating3	0.76	0.66	0.70	0.58	0.46	-					
WI2018 rating1	0.31	0.43	0.43	0.02	0.69	0.24	-				
WI2018 rating2	0.58	0.54	0.30	-0.02	0.42	0.69	0.49	-			
WI2018 rating3	0.76	0.91	0.59	0.17	0.49	0.82	0.27	0.73	-		
GC2018 rating1	0.32	0.75	0.39	0.29	0.33	0.37	0.05	0.05	0.57	-	
GC2018 rating2	0.58	0.89	0.75	0.49	0.66	0.69	0.35	0.32	0.78	0.80	-

Table 5.4 Mean and standard deviation (SD) of downy mildew disease scores (raw data) of 18 ILs and cucumber controls across different years and environments. DM score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.

Materials	GC2018 rating1[a,b]	GC2018 rating2	NC2018 rating1[c]	NC2018 rating2	NC2018 rating3	NC2018 rating4	Categories[d]
9930	6.50 ± 0.53	6.50 ± 0.53	-	-	-	-	S
7120B	0.00 ± 0.00	0.50 ± 0.53	0.5	2.00	3.00	4.00	R
PT0108	0.50 ± 0.53	0.50 ± 0.53	2	2.00	2.00	4.00	R
PT0178	3.00 ± 1.06	4.00 ± 0.00	4	5.00	6.00	7.00	R
PT0214	3.50 ± 2.67	3.50 ± 2.67	-	-	-	-	(seg) R, S
PT0262	0.50 ± 0.53	1.00 ± 0.00	3	5.00	4.00	6.00	R
PT0279	1.00 ± 0.00	-	1	3.00	4.00	5.00	R
PT0301	2.25 ± 2.38	3.25 ± 1.98	3	4.50	5.00	6.00	(seg) R, S
PT0303	5.50 ± 2.07	-	1	4.00	4.00	4.00	S
PT0304	3.13 ± 1.89	3.75 ± 1.91	-	-	-	-	(seg) R, S
PT0307	2.50 ± 0.53	2.50 ± 0.53	3	4.00	6.00	6.00	R
PT0317	3.71 ± 1.70	3.86 ± 2.04	1	1.00	2.00	3.00	(seg) R, S
PT0326	2.75 ± 0.71	-	3.5	4.00	5.50	4.50	I
PT0331	1.50 ± 0.53	3.00 ± 2.14	2	4.00	4.00	6.00	(seg) R, S
PT0332	7.00 ± 0.00	7.50 ± 0.53	5	6.00	7.00	8.00	S
PT0333	6.00 ± 0.00	6.50 ± 0.53	n/a	n/a	n/a	n/a	S
PT0336	3.50 ± 0.53	4.00 ± 1.07	4	6.00	7.00	7.00	I
PT0339	4.50 ± 2.67	5.00 ± 3.21	n/a	n/a	n/a	n/a	(seg) R, S
PT0342	0.5.0 ± 0.53	3.50 ± 0.53	n/a	n/a	n/a	n/a	I
PT0343	1.00 ± 0.00	-	2	2.00	3.00	3.00	R
CCMC	-	-	5	6.00	8.00	8.00	-

Notes:
[a] NC2018 was conducted in Raleigh, NC and GC2018 was conducted in the growth chamber at the University of Wisconsin, Madison.
[b] Disease score ranged from 0 (resistant) to 9 (susceptible).
[c] - meant no seed germination and n/a meant the lines were not grown in NC field due to low seed availability.
[d] Resistant categories were based on growth chamber experiment data. R was resistant, I was intermediate resistant, and S was susceptible.

Table 5.5 Spearman's rank correlation coefficients of ALS disease scores of ILs across different years and environments. Highlighted in grey represent significant correlations (P<0.05).

	GC2018_rating1	GC2018_rating2	NC2018_rating1	NC2018_rating2	NC2018_rating3	NC2018_rating4
GC2018_rating1	-					
GC2018_rating2	0.91	-				
NC2018_rating1	0.59	0.72	-			
NC2018_rating2	0.40	0.62	0.93	-		
NC2018_rating3	0.56	0.69	0.90	0.88	-	
NC2018_rating4	0.49	0.70	0.94	0.95	0.94	-

Table 5.6 List of SSR markers and genotyping results of four ILs for the presence of *dm4.1* and *dm5.1* regions. The marker information was published in Wang et al, (2016).

#	Region	Primer name	PT0108	PT0262	PT0279	PT0343
1	*dm4.1*	UW057915	Gy14	Gy14	Gy14	Gy14
2	*dm4.1*	CsDm4-046	Gy14	Gy14	Gy14	Gy14
3	*dm4.1*	CsDm4-066	Gy14	Gy14	Gy14	GY14+7120B
4	*dm4.1*	CsDm4-047	Gy14	Gy14	Gy14	new pattern
5	*dm4.1*	CsDm4-072	Gy14	Gy14	7120B	GY14+7120B
6	*dm4.1*	CsDm4-071	Gy14	Gy14	7120B	GY14+7120B
7	*dm4.1*	CsDm4-070	Gy14	Gy14	new pattern	failed
8	*dm4.1*	CsDm4-005	Gy14	Gy14	7120B	GY14+7120B
9	*dm4.1*	CsDm4-006	Gy14	Gy14	Gy14	GY14+7120B
10	*dm4.1*	UW057881	Gy14	Gy14	7120B	GY14+7120B
11	*dm4.1*	CsDm4-068	Gy14	Gy14	Gy14	GY14+7120B
12	*dm4.1*	CsDm4-074	Gy14	Gy14	7120B	new pattern
13	*dm4.1*	CsDm4-069	Gy14	Gy14	Gy14	GY14+7120B
14	*dm4.1*	UW004293	Gy14	Gy14	7120B	GY14+7120B
15	*dm4.1*	CsDm4-060	Gy14	Gy14	Gy14	Gy14
1	*dm5.1*	CsDM5-044	Gy14	Gy14	Gy14	Gy14
2	*dm5.1*	CsDM5-063	Gy14	Gy14+7204	GY14+7120B	Gy14 + 9930
3	*dm5.1*	CsDM5-065	Gy14	Gy14	Gy14	Gy14
4	*dm5.1*	CsDM5-066	Gy14	Gy14	new pattern	Gy14
5	*dm5.1*	CsDM5-067	Gy14	Gy14	new pattern	Gy14
6	*dm5.1*	CsDM5-049	Gy14	Gy14	Gy14	Gy14
7	*dm5.1*	CsDM5-068	Gy14	Gy14+7204	Gy14	Gy14+7204
8	*dm5.1*	CsDM5-050	Gy14	Gy14	Gy14	Gy14
9	*dm5.1*	CsDM5-082	Gy14	Gy14	new pattern	Gy14
10	*dm5.1*	CsDM5-086	Gy14	Gy14	new pattern	Gy14
11	*dm5.1*	SSR01280	Gy14	Gy14	new pattern	Gy14
12	*dm5.1*	CsDM5-080	Gy14	Gy14	Gy14	Gy14
13	*dm5.1*	CsDM5-073	Gy14	Gy14	GY14+7120B	Gy14
14	*dm5.1*	CsDM5-074	Gy14	Gy14	new pattern	Gy14
15	*dm5.1*	CsDM5-076	Gy14	Gy14	Gy14	Gy14
16	*dm5.1*	CsDM5-078	Gy14	Gy14	GY14+7120B	Gy14
17	*dm5.1*	UW058914	Gy14	Gy14	Gy14	Gy14

Table 5.7 Mean and standard deviation (SD) of downy mildew disease scores (raw data), genetic background and downy mildew QTL region of four ILs and DM resistant NILs and cucumber controls tested in the growth chamber. DM score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.

Materials	**Disease score**	**Categories**	**Background**	**DM region in NILs and ILs**
7120B	1.25 ± 0.50	R	-	-
Gy14	3.00 ± 0.00	I	-	-
7204	3.50 ± 0.53	S	-	-
9930	5.00 ± 0.00	S	-	-
Z298	3.50 ± 0.53	S	-	-
B5-4	2.13 ± 0.35	R	Gy14	*dm4.1*
E8-5	1.00 ± 0.00	R	Gy14	*dm5.1*
D6-1	2.38 ± 0.52	R	7204	*dm4.1*
H12-16	1.63 ± 0.52	R	7204	*dm5.1*
NIL4-85	3.13 ± 0.35	I	9930	*dm4.1*
NIL5-51	4.50 ± 0.53	S	9930	*dm5.1*
PT0108	1.13 ± 0.35	R	Gy14, Z298	-
PT0262	1.75 ± 0.71	R	Gy14, unknown	-
PT0279	1.63 ± 0.52	R	Gy14, 7204	*dm4.1*
PT0343	2.25 ± 0.46	R	Gy14, 9930, unknown	*dm4.1, dm5.1*

Figure 5.1 Field screening images for angular leaf spot (ALS) resistance in *C. hystrix* TH1 and CN1 accessions and recurrent parent cucumber lines in WI2016 and WI2017 experiments. Disease scores for each line are in parentheses. ALS score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.

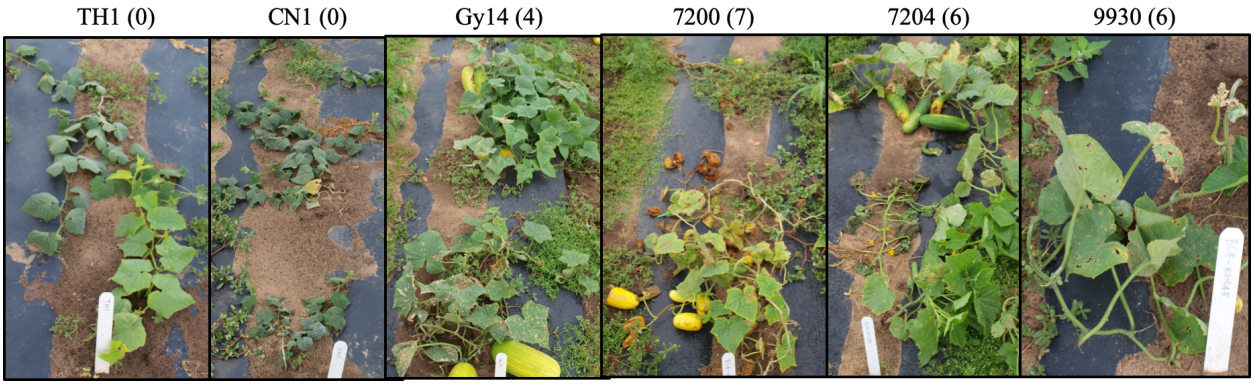


Figure 5.2 Cumulative disease scores for angular leaf spot (ALS) in *C. hystrix* ILs and cucumber resistant line, Gy14, and cucumber susceptible line, 9930, across different environments and rating times. ALS score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.

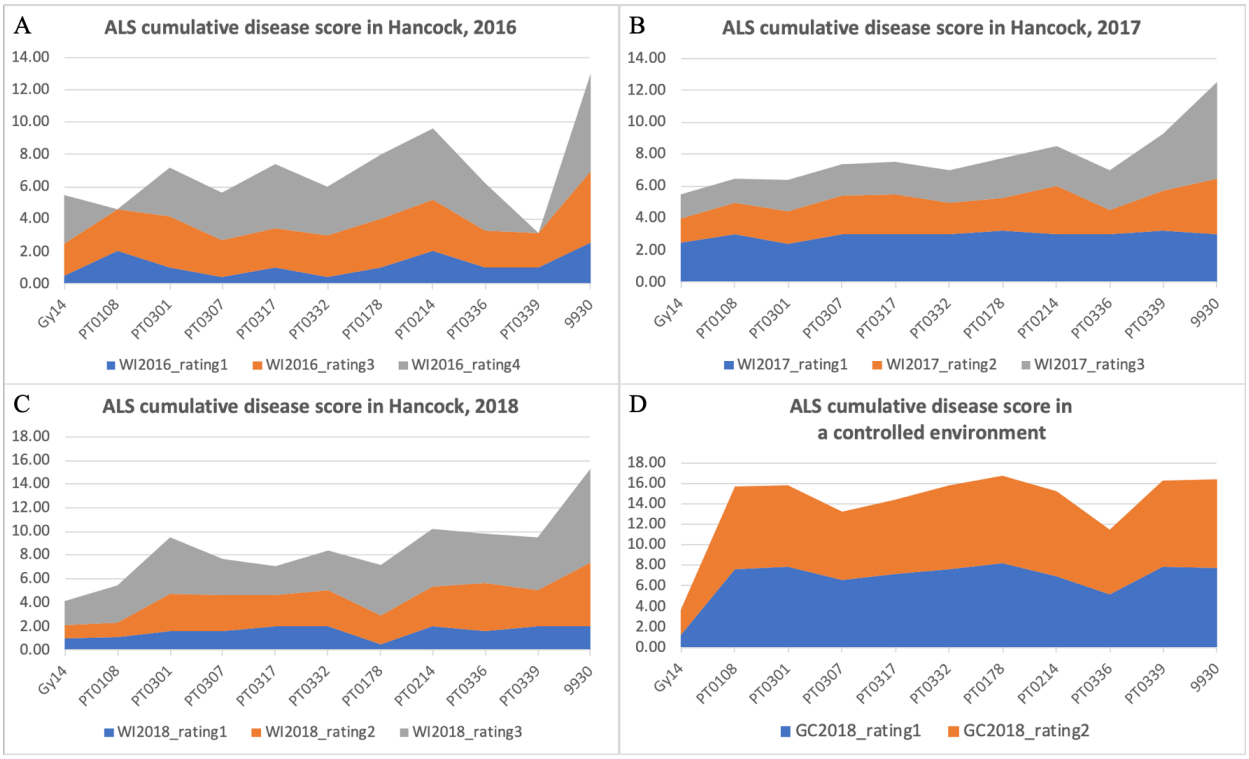
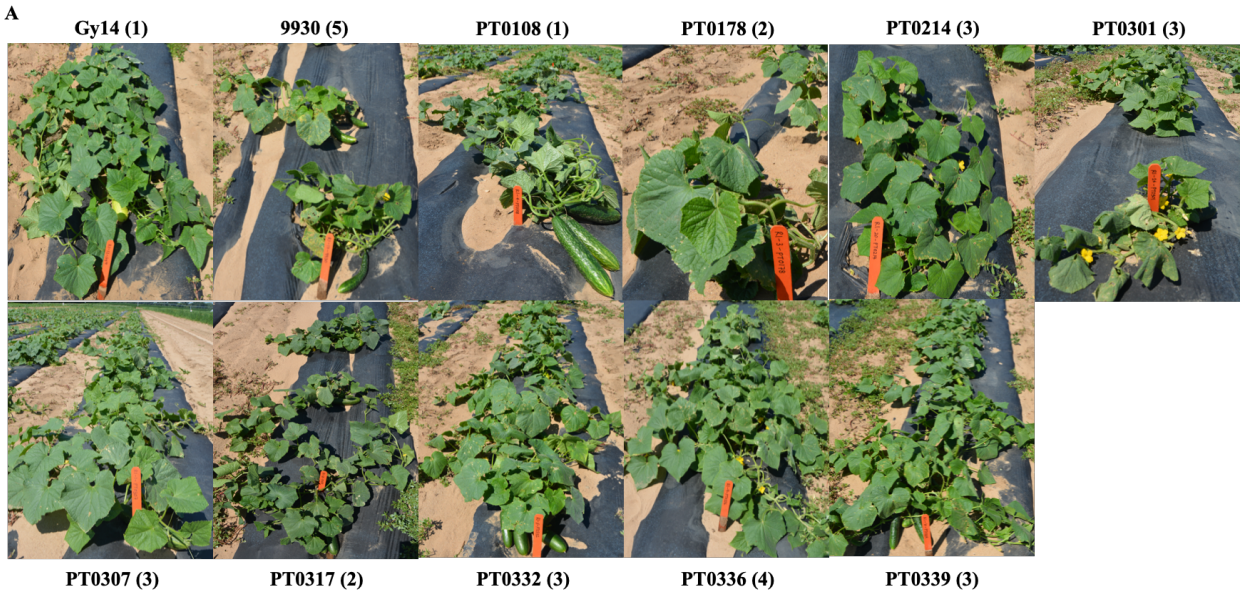


Figure 5.3 Performance of different *C. hystrix* ILs to angular leaf spot resistance evaluated in A) Hancock field (pictures from WI2018_rating2) and B) controlled environment. Disease scores for each line are in parentheses. ALS score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.



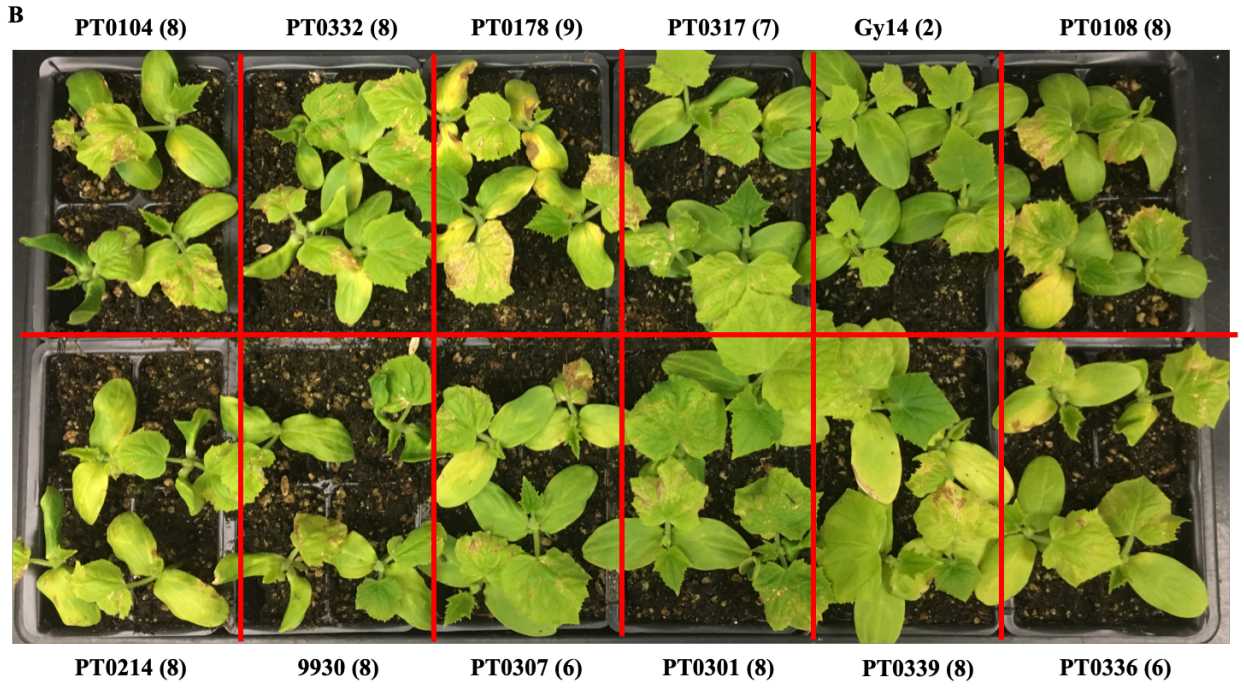


Figure 5.4 Preliminary screening for downy mildew (DM) resistance in *C. hystrix* TH1 and CN1 accessions, an interspecific F_1 hybrid, IS1104, an amphidiploid, $A01S_1$ and recurrent parent cucumber lines. Disease score for each line was in parentheses. DM score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.

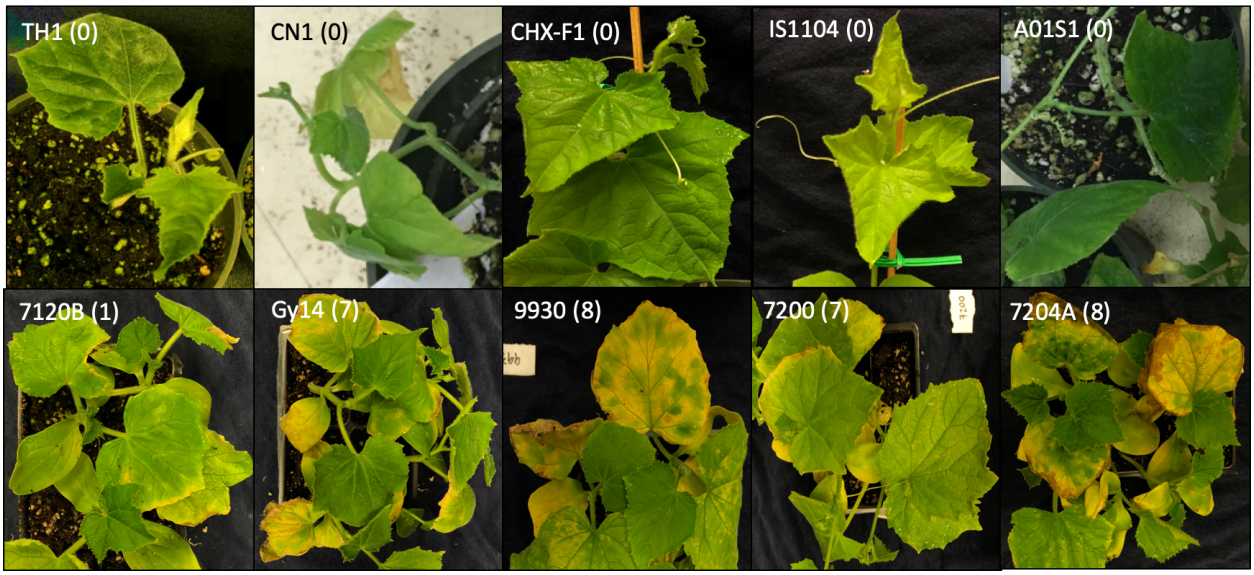


Figure 5.5 Downy mildew disease scores in *C. hystrix* ILs from field experiment in Raleigh, NC 2018 and in the controlled environment. DM score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.

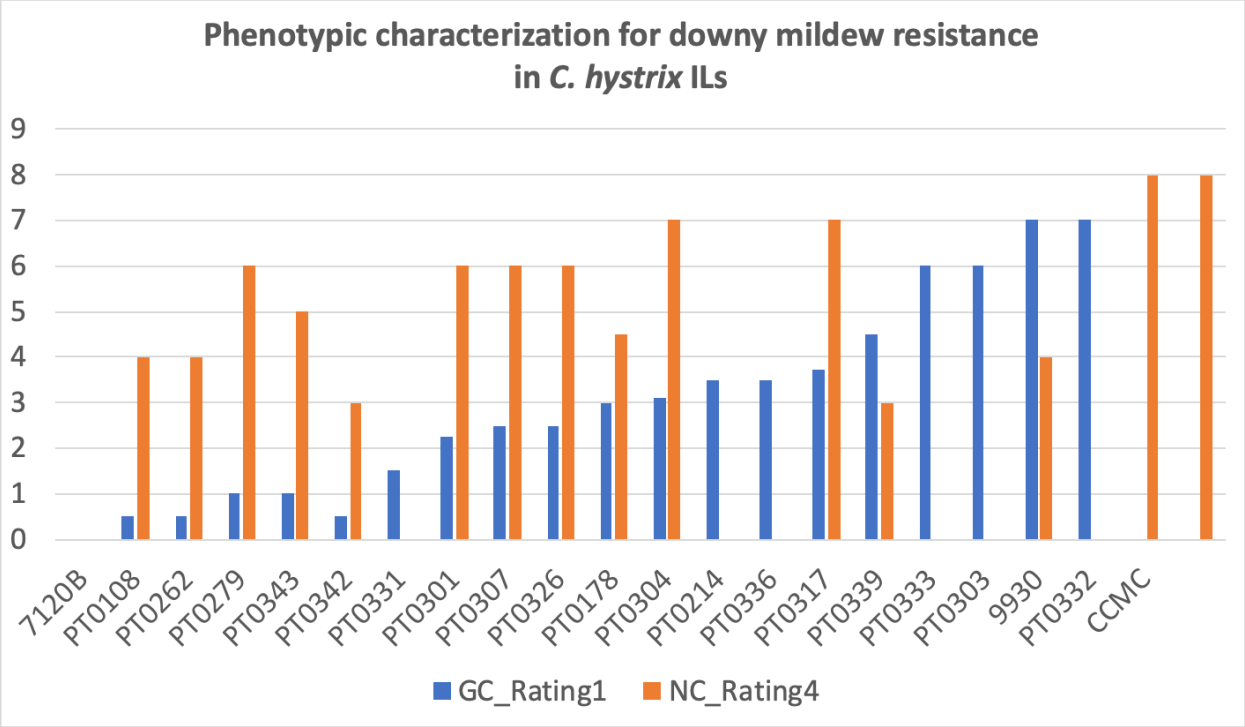


Figure 5.6 Performance of different *C. hystrix* ILs to DM resistant evaluated in the controlled environment. The grand mean disease score for each line was in parentheses. "seg" represented ILs that were segregating for DM resistance trait. DM score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.

Figure 5.7 Downy mildew disease scores in *C. hystrix* ILs that were grown in the controlled environment and evaluated with cucumber of different genetic backgrounds. B5-4, D6-1, and NIL4-85 were introgressed with *dm4.1* in Gy14, 7204 and 9930 background, respectively. E8-5, H12-16, and NIL5-51 were introgressed with *dm5.1* in Gy14, 7204 and 9930 background, respectively. DM score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.

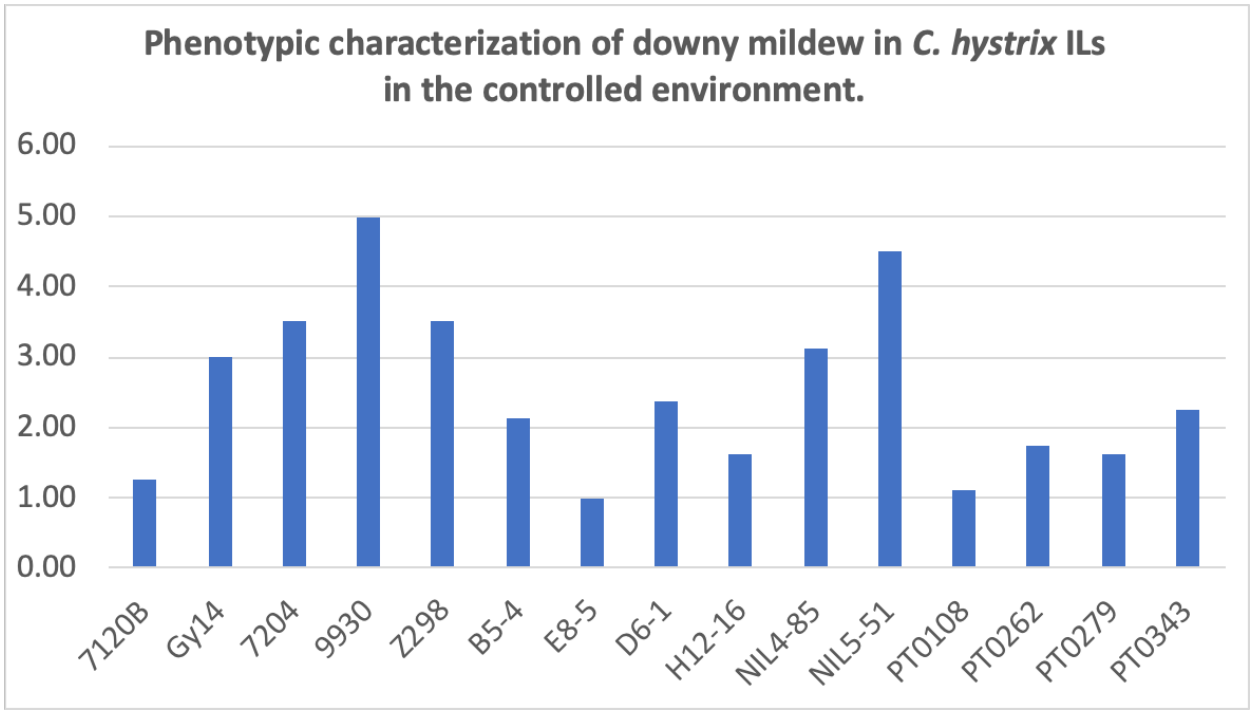


Figure 5.8 Performance of different *C. hystrix* ILs compared to different DM resistant near-isogenic lines (NILs) that had been introgressed with *dm4.1* or *dm5.1* in different cucumber genetic backgrounds. The resistance to DM was evaluated in the controlled environment. The grand mean disease score for each line was in parentheses. DM score was rated on a scale of 0 – 9, where 0 was resistant and 9 was susceptible.

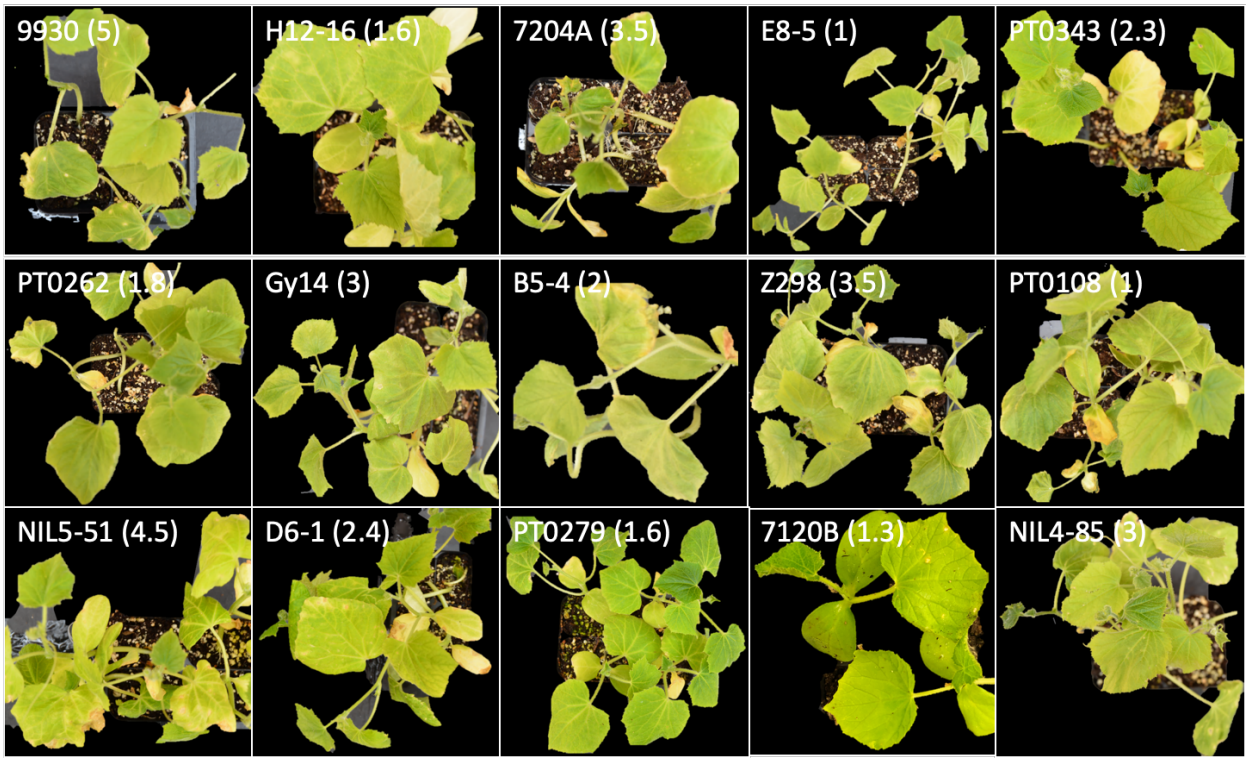


Figure 5.9 The presence of *STAYGREEN* (*CsSGR*) in *C. hystrix*, cucumber and *C. hystrix* ILs. 1) *C. hystrix* TH1, 2) *C. hystrix* CN1, 3) WI2757, 4) Gy14, 5) 7120B, 6) PI7088, 7) Z298, 8)7204, 9) 9930, 10) amphidiploid A01S_1, 11) PT0108, 12) PT0262, 13) PT0279, 14) PT0301, 15) PT0307, 16) PT0326, 17) PT0331, 18) PT0343, 19) PT0332, 20) PT0339, 21) PT0342, 22) PT0178, 23) PT0214, 24) PT0303, 25) PT0304, 26) PT0317, 27) PT0333 and 28) PT0336.

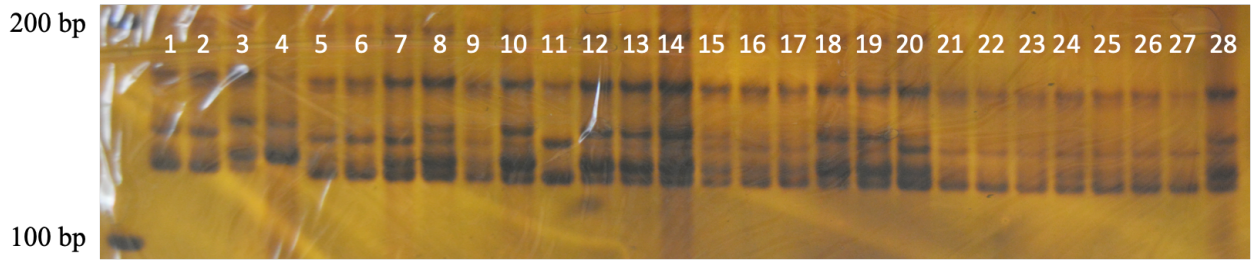


Figure 5.10 Pictures of plants and fruits of *C. hystrix* ILs A-B) PT0108, C-D) PT0262, E-F) PT0279, and G-I) PT0343

Appendix 5.1 Phenotypic data for male and female flowering time, plant structure, and fruit measurements for *C. hystrix* and four ILs collected in Hancock, WI 2016-2018.

The Appendix 5.1 is available at the following link:
https://cranberry.vcru.wisc.edu/nextcloud/index.php/s/xyAxJorLr3j9mJB